VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky

DIPLOMOVÁ PRÁCE

2012 **Jan Paliga**

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra telekomunikační techniky

Uložení dat v diskových polích

Storage of Data in Disks Arrays

2012 **Bc. Jan Paliga**

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra telekomunikační techniky

Zadání diplomové práce

Student: **Bc. Jan Paliga**

Studijní program: N2647 Informační a komunikační technologie

Studijní obor: 2601T013 Telekomunikační technika

Téma: Uložení dat v diskových polích
Storage of Data in Disks Arrays

Zásady pro vypracování:

Pro uložení rozsáhlých dat se používají disková pole, která zároveň poskytují odolnost proti poruše jednoho nebo dvou disků. Princip odolnosti proti poruchám je označován jako RAID. Cílem diplomové práce je prezentovat základní principy RAID a navrhnout laboratorní úlohu.
Diplomová práce bude obsahovat:
1. Přehled ukládání dat pomocí datových polí RAID včetně nastupujících trendů.
2. Sestavení interaktivní prezentace jako studijní materiál ve formátu Flash.
3. Návrh laboratorní úlohy s diskovým polem RAID.

Seznam doporučené odborné literatury:

VADALA, D. *Managing RAID on Linux.* [s.l.] : O'Reilly Media, 2002. 264 s. ISBN 978-1565927308.
PRESTON, W.C. *Using SANs and NAS.* [s.l.] : O'Reilly Media, 2002. 224 s. ISBN-13:978-0-596-001.
ŠTEKL, P. *Testujeme PC s programem Sisoft Sandra.* Brno : Computer Press, 2003. 136 s. ISBN 80-251-0012-X.
LUCAS, M. *FreeBSD : Podrobný průvodce síťovým operačním systémem.* [s.l.] : Computer Press, 2003. 642 s. ISBN 80-7226-795-7.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **doc. Ing. Jaroslav Zdrálek, Ph.D.**

Datum zadání: 19.11.2010
Datum odevzdání: 04.05.2012

VYSOKÁ ŠKOLA BÁŇSKÁ-TECHNICKÁ UNIVERZITA Fakulta elektrotechniky a informatiky OSTRAVA

prof. RNDr. Vladimír Vašinek, CSc.
vedoucí katedry

prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlášení studenta

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

Dne: 4.5.2012 ……

Podpis

Poděkování

Rád bych na tomto místě poděkoval vedoucímu mé diplomové práce doc. Ing. Jaroslavu Zdrálkovi, Ph.D a panu Ing. Adamu Němčekovi za jejich pomoc, podněty k tématu, ochotu a věnovaný čas při zpracovávání této diplomové práce.

Abstrakt

Tato práce se zabývá ukládáním dat do polí RAID a skládá se ze tří hlavních částí:

První část přehledně shrnuje principy a úrovně RAID, obsahuje jejich porovnání a schéma uložení informací jednotlivých úrovní RAID na fyzických pevných discích. Popisuje systémy, v nichž jsou tato pole obsažena a jejich druhy připojení k ostatním systémům.

Druhá část obsahuje popis a návrh interaktivní prezentace ve formátu FLASH jako pomůcku při studiu tématu RAID polí.

Poslední třetí část je návrhem laboratorní úlohy, při níž si studenti ověří v praxi některé vlastnosti a chování diskových polí RAID a seznámí se s ovládáním a konfigurací zástupce NAS zařízení s implementovanou RAID architekturou.

Klíčová slova

RAID, pevný disk, HDD, SSD, pole pevných disků, ethernet, fiber channel, striping, mirroring, FLASH, propustnost, spolehlivost, redundance, CRC, FC, SCSI, ATA, USB, HBA, NAS, DAS, SAN, SPARE, Thecus N5200B, základní RAID úrovně, hybridní RAID úrovně, DDF

Abstract

This paper deals with the storing of data to the RAID arrays and consists of three main parts:

The first section summarizes the principles and RAID levels clearly, includes the schema of stored information on physical hard drives at different levels of RAID. It describes systems in which these fields are included and the types of connections to other systems.

The second part describes the design and description of an interactive Flash presentation as an aid in studying the topic of RAID arrays.

The third and the last part deals with the design of a laboratory task in which students examine in practice some of the properties and behavior of RAID and get familiar with the operation and configuration of a NAS device with RAID architecture implemented.

Keywords

RAID, a hard drive, HDD, SSD, an array of the hard drives, ethernet, fiber channel, striping, mirroring, FLASH, throughput, reliability, redundancy, CRC, FC, SCSI, ATA, USB, HBA, NAS, DAS, SAN, SPARE, Thecus N5200B, standard RAID levels, hybrid RAID levels, DDF

Seznam použitých symbolů

μ	[počet oprav/jedn. času]	intenzita opravy
λ	[počet poruch/jedn. času]	intenzita poruchy
C	[kus]	počet paritních disků ve skupině
D	[kus]	počet datových disků ve skupině
G	[kus]	počet skupin v poli
n	[kus]	počet datových disků v poli
ND	[kus]	počet disků
n_G	[kus]	počet datových disků ve skupině
m	[kus]	počet disků ve vnořeném poli
r	[%/100]	chybovost
X	[jedn. informace/s]	datová propustnost jednoho disku

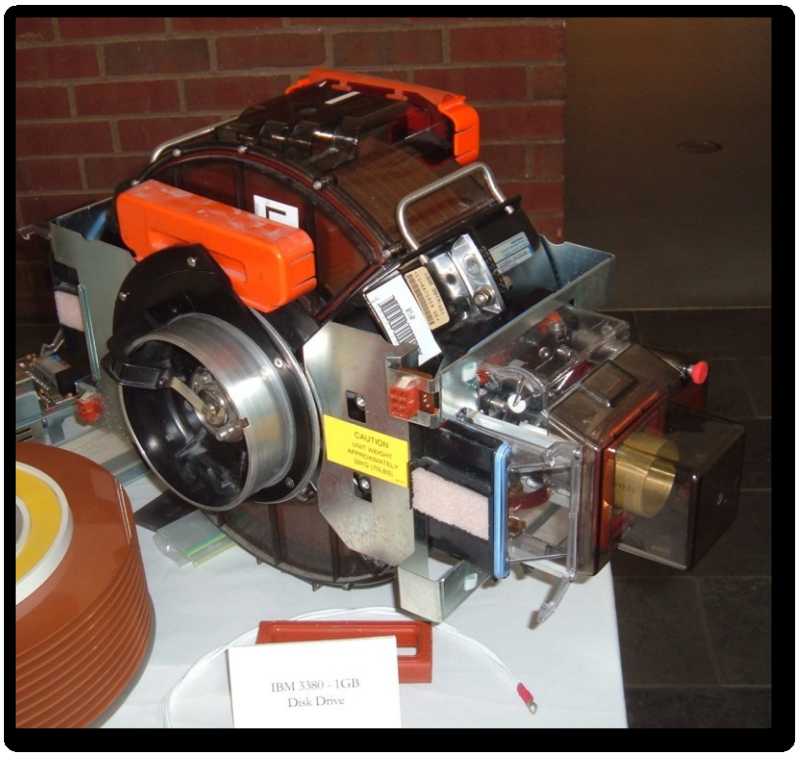
Seznam použitých zkratek

AAM (Automatic acoustic management)	technologie umožňující regulaci hluku ATA zařízení
ANSI (American National Standards Institute)	Státní úřad pro vydávání standardů v USA
AT, IBM/AT (Advanced Technology)	druhá generace architektury společnosti IBM zveřejněná v roce 1984
ATA (AT Attachment)	rozhraní pro připojení paměti v architekturách IBM/AT
ATAPI (ATA Packet Interface)	protokol rozšiřující ATA interface o možnost přenášet SCSI příkazy
Array Roaming	vlastnost diskových polí, umožňující provoz diskového setu na jiném stroji bez rozpadu pole
DAS (Direct Attached Storage)	koncepce datového úložiště, ve které je úložiště připojeno přímo k serveru
DCO (Device Configuration Overlay)	část datové oblasti disku nepřístupné operačnímu systému
DDF (Disk Data Format)	analogie souborového systému používaného v polích RAID
ENTERPRISE	zařízení určená do velkých podnikových systémů
FC-AL (Fibre Channel Arbitrated loop)	sériový transportní protokol a topologie využívající příkazy SCSI
FLASH	elektricky programovatelná paměť, softwarová technologie spol. ADOBE
HBA (Host Bus Adapter)	připojuje další zařízení nebo síť k hostitelskému systému
HPA (Host Protected Area)	část datové oblasti disku nepřístupné operačnímu systému
HHD (Hybrid Hard Drive)	paměťové zařízení založené na mechanickém pohybu a elektricky ovládané semipermanentní paměti
MTTDL (Mean Time To Data Lost)	střední doba do ztráty dat

MTTF (Mean Time To Failure)	střední doba do poruchy
MTTR (Mean Time To Repair)	střední doba opravy
PATA (Paralel ATA)	paralelní verze ATA rozhraní (dříve také jen ATA, nebo IDE)
SAN (Storage Area Network)	koncepce datového úložiště, ve které jsou datové sklady připojeny do dedikované sítě
SAS (Serial Attached SCSI)	sériová sběrnice využívající příkazy SCSI
SATA (Serial ATA)	zkratka pro sériovou (novější) verzi ATA
SCSI (Small Computer System Interface)	sada komunikačních příkazů a rozhranní mezi sběrnicí počítače a periferiemi
S.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology)	systém monitoruje a analyzuje „zdraví“ paměťové jednotky
SOHO (Small Office / Home Office)	zařízení určená pro provoz v domácích a malých kancelářích
SPARE	záložní disk, který je automaticky po selhání jiného disku nasazen
SSA (Serial Storage Architecture)	sériový transportní protokol využívající příkazy SCSI
ZFS (Zettabyte File System)	souborový systém odolný proti chybám
Z-RAID, Z-RAID2	architektura úložiště založená na bázi ZFS

Obsah

1. Úvod



Obr. 1: Pevný disk IBM 3380

Cílem první části tohoto dokumentu, je vážené čtenáře blíže seznámit se současnými principy ukládání dat do polí RAID.

Jako první se podíváme trochu do historie a objasníme si okolnosti vzniku diskových polí. Definujeme si systémy odolné proti poruchám a objasníme jejich dělení. Pokračovat budeme popisem komponent, ze kterých se pole skládá a jak jsou jednotlivé části propojeny do systému zpracovávající data z těchto úložišť. Vysvětlím také principy tří hlavních typů úložišť a opět popíši připojení úložiště k systémům resp. cestu dat k uživateli.

V textu se dále pokusím shrnout všechny používané úrovně, popsat jejich architekturu, kladné a záporné vlastnosti podrobně rozeberu použité technologie s přihlédnutím k jejich použití do budoucnosti. Podrobněji popíši důležité termíny, používané v této oblasti jako jsou parita, spolehlivost, mirroring, striping aj.

Pozornost budu věnovat také hybridním úrovním polí RAID, jednotlivá vnoření znázorním graficky. Na konci této části se budu naopak věnovat pohledu do budoucnosti a nastupujícím technologiím a architekturám.

V druhé části textu popíši tvorbu a ovládání prezentace ve formátu FLASH. Součástí citované aplikace bude i průvodce laboratorní úlohou popsanou níže. Na závěr části přiložím zdrojový kód v jazyce Action Script 3.0 a ukázku grafického zpracování.

Úkolem závěrečné části dokumentu je návrh laboratorní úlohy, která má studenty seznámit s touto tématikou prostřednictvím praktické ukázky využití NAS systému. K tomuto

účelu mi byl zapůjčen SOHO NAS system Thecus N5200B osazený pěti disky SATA o celkové kapacitě 2,5TB.

1.1. RAID (Redundant Arrays of Inexpensive Disks)

1.1.1. Důvod a vznik

V devadesátých letech dvacátého století zaznamenal růst výkonu výpočetních systémů nebývalý nárůst. Výpočetní systémy se začaly stávat nevyváženými, co do poměru maximální výkonnosti základní procesorové jednotky a maximálních vstupně výstupních operací. Podle Amdahlova zákona [2] bude aplikace běžící na stroji stokrát rychlejším pouze o necelých deset procent rychlejší v důsledku stejně rychlých vstupně výstupních operací, na jejichž vybavení bude čekat devadesát procent svého času.

Určitou nedokonalost mechanických paměťových jednotek, chcete-li - pevných disků, částečně eliminovala zvětšující se kapacita primárních pamětí první úrovně založených na principu elektricky ovládané paměťové buňky, tj. bez mechanických součástí. Vývoj pamětí tohoto typu šel ruku v ruce s vývojem základních procesorových jednotek a i po nákladové stránce zaznamenával podobný trend. Bohužel levné paměti větších kapacit nebyly řešením pro všechny druhy aplikací jako například aplikací zpracovávajících velké množství multimediálních kontinuálních dat.

V těchto a podobných případech ztrácí použití relativně malých vyrovnávacích pamětí smysl a rychlost přísunu dat ke zpracování a rychlost ukládání již zpracovaných dat závisí z větší části na paměťových jednotkách. Pro tyto mechanické součástky jsou určujícími parametry vystavovací doba, doba jedné otáčky a hustota záznamu. V průběhu deseti let 1971 až 1981 se vystavovací čas u high-end disků IBM zlepšil pouze dvakrát, zatímco čas potřebný na jednu otáčku se nezměnil. Větší hustota záznamu znamená větší přenosovou rychlost, více čtecích hlav má schopnost snížit průměrnou vystavovací dobu, ale fyzická vystavovací doba se u vyvíjených pevných disků zvyšuje meziročně pouze o 7%. Není proto důvodné se domnívat, že tomu bude jinak i v letech příštích, ba co více, při zachování základních funkčních principů se pomalu blížíme k fyzikálním hranicím těchto datových médií.

Výše uvedený fakt vedl výzkumníky k hledání řešení jiného, než nasazování nákladných profesionálních pevných disků do systémů se zlomkovou hodnotou. V roce 1987 proto výzkumníci z University of California - Berkeley u příležitosti International Conference on Management of Data definovali poprvé zkratku RAID jako pole redundantních levných disků a představili základní úrovně jejich řazení. Jak již název napovídá, šlo především o řešení

problému dostupnosti vyšších kapacit úložišť při zachování nízké ceny za paměťovou jednotku. Autoři citovaného dokumentu *lit. [1]* udávají jako jeden z referenčních pevný disk společnosti IBM konkrétně typ 3380 model AK4 *vi.z Obr. 1* určený pro využití v mainframe počítačích, jež v roce vydání dokumentu reprezentoval 70 000USD - 120 000USD podle konfigurace a kapacity. Autoři se snaží vhodným řazením menších, pomalejších a logicky i levnějších pevných disků docílit podobných parametrů.

1.1.2. Současný stav a klasifikace

Následující léta nepřinášela z pohledu technologií nic převratného a vyznačovala se především pozvolným navyšováním kapacit disků a zmenšováním rozměrů tj. zvětšováním hustoty záznamu na jednotku plochy *viz.Obr. 2* .

Obr. 2:Zástupce současných pevných disků

Způsob řazení disků do polí zůstává nezměněn, pouze je původních pět RAID úrovní, definovaným v *lit. [1],* postupně rozšířeno o další dvě úrovně. Postupem času vznikají také další proprietální řešení výrobců hardwaru a značení jednotlivých úrovní začíná být nepřehledné. Z takzvaných vložených (nested), dříve hybridních, úrovní označovaných rekurentně číslem úrovně vnořené, znaménkem plus (+) a úrovně nadřazené (například 1+0, 5+0 apod.) vypadlo spojovací znaménko a dále jsou uváděny pouze jako spojení dvou číslovek (například 10,50 apod.). Zmiňovaný fakt vedl skupinu Raid Advisory Board k vydání doporučujícího dokumentu *lit. [3],* kde nově definují tři základní skupiny úrovní RAID konfigurací:

A) Systémy resistentní (FRDS) musí splňovat kritéria:

- Odolnost proti ztrátě dat a přístupu k datům v souvislosti s poruchou disku
- Rekonstrukci obsahu disku v poruše na nový disk
- Odolnost proti ztrátě dat v důsledku "write hole"

- Odolnost proti ztrátě dat v důsledku selhání hostitele a jeho vstupně-výstupní sběrnice
- Odolnost proti ztrátě dat v důsledku selhání výměnné jednotky
- Monitoring a indikaci selhání výměnné jednotky

B) Systémy odolné proti chybě (FTDS) musí splňovat tato kritéria:

- Automatický diskový "swap" a "hot swap"
- Odolnost proti ztrátě dat v důsledku selhání vyrovnávací paměti
- Odolnost proti ztrátě dat v důsledku výpadku napájení
- Odolnost proti ztrátě dat v důsledku nedodržení teplotního rozsahu
- Varování nefunkční výměnné jednotky a chyby prostředí
- Odolnost proti ztrátě přístupu k datům v důsledku selhání kanálu
- Odolnost proti ztrátě přístupu k datům v důsledku selhání modulu řadiče
- Odolnost proti ztrátě přístupu k datům v důsledku selhání vyrovnávací paměti
- Odolnost proti ztrátě přístupu k datům v důsledku selhání napájecího zdroje

C) Systémy odolné proti katastrofě (DTDS) musí splňovat tato kritéria:

- Odolnost proti ztrátě přístupu k datům v důsledku selhání hostitele a jeho vstupně-výstupní sběrnice
- Odolnost proti ztrátě přístupu k datům v důsledku selhání napájecího zdroje
- Odolnost proti ztrátě přístupu k datům v důsledku výměny komponentu
- Odolnost proti ztrátě dat a přístupu k datům v důsledku selhání více disků
- Odolnost proti ztrátě přístupu k datům v důsledku selhání zóny
- Odolnost proti ztrátě dat v důsledku selhání vzdálené zóny

Původní (Berkley) klasifikace zůstává i nadále v platnosti, jelikož nová definice tříd s ní není nijak v rozporu.

1.1.3. Spolehlivost RAID

Spolehlivost je vlastnost systémů, kvůli které jsou mimo jiné pole RAID tak oblíbené a vyhledávané. Spolehlivost systémů vždy závisí na spolehlivosti jednotlivých součástí, ze kterých je celý systém složen.

Jedním z parametrů, které nás informují o spolehlivosti zařízení je doba bezporuchového provozu systému. V oblasti diskových polí, a nejen tam, se častěji používá jeho převrácená hodnota, což je doba do poruchy MTTF.

Za předpokladu stejné doby do poruchy u všech disků v poli, spolehlivost diskového pole neodolného vůči chybám logicky klesá s počtem disků, v dotyčném poli obsažených, podle vztahu (*1.1)*.

$$MTTF_{DA} = \frac{MTTF_{SD}}{ND} \tag{1.1}$$

Kde

$MTTF_{DA}$ - střední doba do poruchy diskového pole,(Mean Time To Failure of disk arrea)

$MTTF_{SD}$ - střední doba do poruchy samostatného disku, (Mean Time To Failure of a single disk)

ND – počet disků (Number of Disks)

Výpočet MTTF pro redundantní úrovně RAID systémů (*1.2)*:

$$MTTF_{DA} = \frac{(MTTF_{SD})^2}{(D + C \cdot n_G) \cdot (G + C - 1) \cdot MTTR} \tag{1.2}$$

Kde:

$MTTF_{DA}$ - střední doba do poruchy diskového pole,(Mean Time To Failure of disk area)

$MTTF_{SD}$ - střední doba do poruchy samostatného disku, (Mean Time To Failure of a single disk)

MTTR - střední doba opravy systému, (Mean Time To Repair)

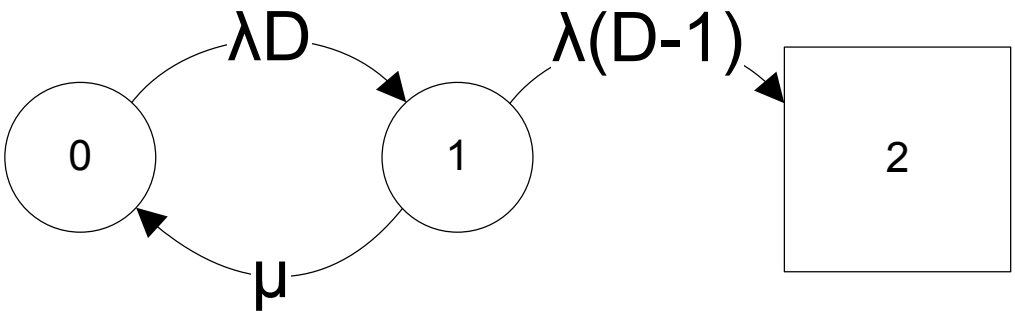
G - počet datových disků ve skupině (number of disk in a Group)

D - počet datových disků (number of Data disks)

n_G =D/G – počet skupin (number of group)

C - počet kontrolních disků ve skupině (number of Check disk in a group)

Chování RAID polí v různých stavech nejlépe znázorníme pomocí Markovových modelů. Pro úrovně s jednoduchou distribuovanou paritou platí *Obr. 3:*



Obr. 3:Markovovův model pro pole s distribuovanou jednoduchou paritou

Kde:

λ - intenzita poruchy

μ - intenzita opravy

D - počet disků

0 - výchozí stav systému se všemi funkčními disky

1 - stav systému po selhání jednoho disku – systém v degradovaném režimu

2 - systém nefunkční – ztráta dat

Spolehlivost je natolik komplexní veličina resp. parametr, že takto vyjádřená nám ještě zcela neposkytuje přesný obraz o spolehlivosti systému. Parametry MTTF nebo MTTDL nám poskytují pouze jakési vodítko a představu o spolehlivosti daného systému. Spolehlivost ovlivňuje celá řada dalších vlivů, mezi něž patří například systémové havárie, neopravitelné bitové chyby, korelační diskové chyby apod.

Ke zvyšování spolehlivosti velkou měrou přispívají právě redundantní systémy, ať se jedná o redundance celých systémů, které jsou třeba i geograficky vzdálené z důvodu případné lokální katastrofy nebo o ty základní podsystémy, ze kterých jsou pak složeny složité celky s vysokou dostupností. Pojďme si nyní představit nejrozšířenější zástupce systémů pro ukládání dat odolných proti chybám a jejich součásti.

2. HBA (Host Bus Adapter)

HBA je komponenta systému, známá také jako řadič nebo adapter, která vytváří interface-rozhraní mezi sběrnicí hostitelského systému a dalším systémem jako například pamětí nebo sítí. V širším slova smyslu je HBA obsaženo ve všech typech datových úložišť.

HBA fyzicky většinou tvoří obvod, ať už integrovaný do základní desky nebo na zvláštním modulu resp. desce plošných spojů, který převádí a adaptuje signály ze systémové sběrnice na signály daného systému, jež má být připojen. Typickým příkladem HBA jsou SCSI, SATA, IDE, ethernetové nebo optické síťové adaptéry.

3. Typy rozhraní

Zařízení pro ukládání dat jsou v největší míře, odhlédneme-li od proprietárních řešení a málo rozšířených rozhraní, připojována k HBA jednoho z klonů ATA nebo SCSI interface. Jelikož jsou využívána v současné době a pravděpodobně ještě nějakou dobu budou, popíšeme si je v průběhu této kapitoly podrobněji.

3.1. SCSI (Small Computer System Interface)

SCSI je soubor ANSI standardů umožňující komunikaci mezi HBA a periferiemi jako například pevnými disky, skenery, zálohovacími zařízeními a podobně. Existují v několika verzích lišících se mimo jiné propustností a spolehlivostí. V rámci SCSI sběrnice jsou definovány standardní rozhraní a sada příkazů pro komunikaci mezi periferií a řadičem sběrnice, ke které je periferie připojena.

3.1.1. Paralelní SCSI

Je základní variantou standartu SCSI a tudíž nejstarší. V současné době bylo nahrazeno sériovou podobou. Existovalo v několika verzích a rychlostech *viz Tab. 1;SCSITA,2011.*

Rozhraní	Alias	Šířka [bit]	Časování [MHz]	Maximální		
				propustnost [MB/s]	délka [m]	počet zařízení
SCSI-1	Narrow SCSI	8	5	5	25	8
Fast SCSI		8	10	10	25	8
Fast-Wide SCSI		16	10	20	25	16
Ultra SCSI	Fast-20	8	20	20	25	8
Ultra Wide SCSI		16	20	40	25	16
Ultra2 SCSI	Fast-40	8	40	40	25	8
Ultra2 Wide SCSI		16	40	80	25	16
Ultra3 SCSI	Ultra-160; Fast-80 wide	16	40 (**)	160	12	16
Ultra-320 SCSI	Ultra-4; Fast-160	16	80 (**)	320	12	16
Ultra-640 SCSI (*)	Ultra-5	16	160 (**)	640	(*)	16

Tab. 1:Přehled paralelních SCSI variant lit. [5]

(*)existuje pouze vývojová verze – bude nahrazena sériovou verzí

(**)Double Data Rate – sběrnice přenáší data při sestupné i vzestupné hraně hodin

3.1.2. Sériové SCSI

Tento typ SCSI zachovává z větší části příkazy verze paralelní, avšak šíře jeho sběrnice je jeden bit – sériový přenos. Při přechodu na sériový protokol se vývoj ubíral třemi hlavními směry a to SSA, FC-AL a SAS. Opět existuje řada verzí se vzrůstající rychlostí v závislosti na době vzniku, viz stručná geneze v tabulce *Tab. 2.*

Rozhraní	Šířka [bit]	Časování [GHz]	Maximální		
			propustnost [MB/s]	délka [km]	počet zařízení
SSA	1	0,2	40	0,025	96
SSA 40	1	0,4	80	0,025	96
FC-AL 1Gb	1	1	100	0,5/3(*)	127
FC-AL 2Gb	1	2	200	0,5/3(*)	127
FC-AL 4Gb	1	4	400	0,5/3(*)	127
Serial Attached SCSI (SAS)	1	3	300	0,006	256
iSCSI	Závislé na použité síti				

Tab. 2: Přehled sériových standardů s implementací SCSI lit. [6]
(*) 500 metrů pro multi-mode, 3 kilometry pro single-mode optická vlákna

3.1.1. iSCSI

iSCSI je technologie umožňující přenášet standardní sadu SCSI příkazů přes IP sítě jako například Internet a tím propojit dvě zařízení na velkou vzdálenost bez použití vlastní dedikované kabeláže nebo infrastruktury.

3.2. ATA (AT Attachment)

Další skupinou standardů, s níž je možné se setkat, je ATA – AT Attachment. Standardy ATA spravuje v rámci ANSI skupina INCITS [4]. K přenosu signálu se na této sběrnici používali ATA příkazy určené pro komunikaci výhradně s pevnými disky. Později ovšem vyvstala potřeba připojovat k ATA také například výměnná média a tak bylo nutné sadu příkazů rozšířit. Výsledkem byl ATAPI – ATA Packet Interface protokol, který byl ATA kompatibilní a navíc obsahoval rozšíření v podobě příkazů SCSI přenášených po ATA sběrnici.Stejně jako je tomu u SCSI, i zde jsou již k dispozici dvě topologie a to paralelní a sériová s tím, že se zcela přechází na sběrnici sériovou.

3.2.1. PATA

Jelikož je, jak už bylo zmíněno, paralelní verze ATA na ústupu, ale stále se s ní můžeme setkat, uvedeme si zde jen ve stručnosti její vývoj *viz Tab. 3.*

Rozhraní	Alias	Maximální propustnost [MB/s]	Maximální velikost HDD (512B/sektor)	Nové vlastnosti
IDE (pre-ATA)	IDE	3,3	2,1 GB	22b LBA
ATA-1	ATA, IDE	16,7	137 GB	28b LBA
ATA-2	EIDE, Fast ATA, Ultra ATA	16,7	137 GB	PCMCIA konektor,příkaz Identify Drive
ATA-3	EIDE	16,7	137 GB	S.M.A.R.T., 44 pinový konektor pro 2.5" disky
ATA/ATAPI-4	ATA-4, Ultra ATA/33	33,3	137 GB	podpora ATAPI, HPA,CFA – podpora pro SSD
ATA/ATAPI-5	ATA-5, Ultra ATA/66	66,7	137 GB	kabel - 80 žil; CompactFlash konektor
ATA/ATAPI-6	ATA-6,Ultra ATA/100	100	144 PB	48b LBA, DCO, AAM
ATA/ATAPI-7	ATA-7,Ultra ATA/133	133	144 PB	SATA 1.0, příkazy pro souvislá data, příkazy pro dlouhé logické/fyzické sektory
ATA/ATAPI-8	ATA-8	600	144 PB	podpora pro HHD

Tab. 3: Přehled variant PATA lit.[7]

3.2.2. SATA

Tento v současné době nejrozšířenější klon ATA standardu využívají především pracovní stanice, ale vzhledem ke svým vlastnostem ho můžeme nalézt i v levnějších serverových řešeních. Podporuje tzv. Hot-Plug – připojení za chodu a technologii NCQ - Native Command Queuing, která optimalizuje pohyb čtecích hlav v závislosti na umístění dat. Obě nové vlastnosti umožňují nasazení v poloprofesionální oblasti. Hot-Plug se využívá v osazení RAID polích SATA disky, kdy je lze měnit za chodu a zvyšuje tímto Up-Time pole. Využitím NCQ lze docílit většího výkonu diskového pole resp.

samostatného disku pro Multi Tasking operačních systémů nebo pro aplikace užívající ke své činnosti více fragmentovaná data. SATA existuje v několika revizích a variantách *viz Tab. 4.*

3.2.3. eSATA

Pro připojení externích zařízení je určena varianta eSATA, která ovšem nedisponuje přenosovou rychlostí SATA revize 3.0, ale zato prodlužuje maximální délku datového kabelu a používá robustnější konektor.

3.2.4. eSATAp

Ačkoliv každá moderní základní deska obsahuje konektor eSATA, není jeho používání nikterak samozřejmé, jako je tomu například u rozhraní USB. Z větší části je toto způsobeno malou univerzálností tohoto typu SATA, jelikož mu chybí důležitá vlastnost a to napájení externího zařízení, které je pak nutno napájet z jiného zdroje. Nutnost napájet externě vede k malé rozšířenosti eSATA a neochotě výrobců jej používat v nových zařízeních. Neduh absence napájení řeší napájené eSATA, které je uváděno pod názvem eSATAp, kde „p“ značí „powered“ z anglického slova napájený. eSATAp je kompatibilní s USB ve všech současných revizích, je rychlejší než USB rev.3 a poskytuje napájení.

Standard	Alias	Propustnost [MB/s]	Frekvence	Maximální délka [m]	Počet zařízení na kanál
SATA 1,5 Gb/s	SATA rev.1.0	150 MB/s	1,5 GHz	1	1
SATA 3 Gb/s	SATA rev.2.0	300 MB/s	3 GHz	1	15(*)
SATA 6 Gb/s	SATA rev.3.0	600 MB/s	6 GHz	1	15(*)
eSATA	eSATA	300 MB/s	3 GHz	2	15(*)
eSATAp	eSATAp	300 MB/s	3 GHz	2	15(*)

Tab. 4 : Přehled variant SATA lit.[8]
(*) při použití multiplikátoru portů

V době vzniku dokumentu pracovala skupina SATA-IO na ratifikaci standardu SATA revision 3.1, který by měl mimo jiné zakotvovat a

specifikovat

- mSATA tj. verzi SATA určenou pro mobilní zařízení,
- USM - Universal Storage Module tj. standard pro ukládání dat ve spotřební elektronice s velkou interoperabilitou využívající poslední revizi SATA s propustností 6Gb/s *lit. [9]*
- eSATAp s propustností 6Gb/s
- a dále pak několik vylepšení a rozšíření v oblasti správy napájení zařízeních připojených přes SATA, lepší podpory disků SSD – Solid State Disk apod.

3.3. FC (Fiber Channel)

V segmentu Enterprise řešení je ponejvíce zastoupena pro propojování úložišť technologie FC – Fibre Channel. Původně vyvinutá technologie pro superpočítače a mainframy se v současné době používá nejvíce na propojení resp. připojení zařízení SAN resp. NAS. Navzdory názvu lze protokolem FC – FCP přenášet data i po metalickém vedení. Při aplikaci tří možných topologií a jejich kombinací, ve kterých lze FC provozovat, je použito více než patnáct různých druhů portů se specifickým použitím.

FC se nedrží sedmi úrovní OSI modelu a vystačí si pouze s pěti vrstvami. Stejně jako technologie popsané výše v tomto dokumentu se i FC během času vyvíjí, což zachycuje následující tabulka *Tab. 5:*

Standard	***Propustnost [MB/s]***	***Maximální délka Multimode [m]***	***Maximální délka Singlemode [m]***	***Doba vzniku***
1GFC	*200*	*0,5-860*	*0,5-5000*	*1997*
2GFC	*400*	*0,5-860*	*0,5-5000*	*2001*
4GFC	*800*	*0,5-860*	*0,5-5000*	*2005*
8GFC	*1600*	*0,5-860*	*0,5-5000*	*2008*

10GFC Serial	*2550*	*0,5-860*	*0,5-5000*	*2004*
16GFC	*3200*	*0,5-860*	*0,5-5000*	*2011*
20GFC	*5100*	*0,5-860*	*0,5-5000*	*?*

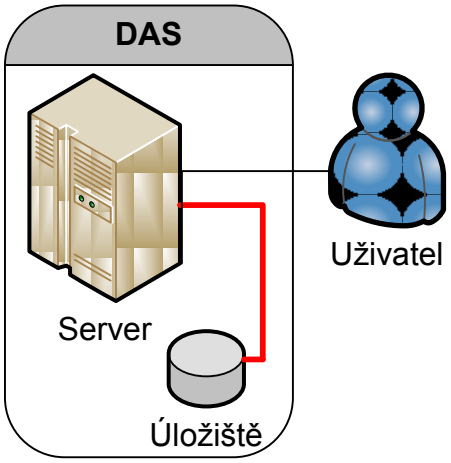
Tab. 5: Přehled variant FC lit. [10]

4. KONCEPCE DATOVÝCH ÚLOŽIŠŤ

V této kapitole bychom si přiblížili jednotlivé koncepce uchovávání dat z hlediska rozprostření datových médií ve fyzickém prostoru a distribuce dat na nich uložených do zpracovávajícího systému nebo ke klientovi. Tyto koncepce jsou v zásadě dvě.

4.1. Lokální - Direct Attached Storage (DAS)

V tomto typu datového úložiště jsou datová média připojeny k řídícímu serveru přes HBA přímo. Zmiňovanou topologii *viz Obr. 4* nazýváme DAS - Direct Attached Storage. Jednotlivá datová média jsou umístěna přímo v serveru nebo v externím diskovém poli poblíž řídícího serveru. Každý server pak využívá svůj datový sklad. Bezprostřední připojení paměťového řadiče k systémové sběrnici poskytuje přístup k jednotlivým fyzickým blokům dat tzv. clusterům.



Obr. 4: Direct Attached Storage

4.1.1. Efektivita a výhody

Tento typ datového úložiště byl prvním a vycházel z konceptu klasického IBM/AT kompatibilního počítače jako serveru. DAS vyniká hlavně svojí jednoduchostí a s tím spojenou dostupností. Topologie tohoto typu je na ústupu pod tlakem zvyšujících se nároků na kapacitu úložišť a s tím úzce související robustnost, bezpečnost, škálovatelnost a správu, jež je při použití přímo připojených datových médií ke každému serveru značně problematická. Vzhledem k jednoduchosti a nákladům spojených s implementací DAS je zmiňovaná topologie

ještě stále hojně rozšířena. Jednoduchou variantu DAS získáme vlastně již sdílením diskového prostoru svého osobního počítače nebo notebooku. Musíme si být však vědomi zranitelností, subtilitou a omezeným poskytovaným výkonem našeho řešení.

4.1.2. Omezení a nevýhody

Limitujícím faktorem je mimo jiné fakt, že pokud mají klienti pracovat s poskytovaným diskovým prostorem, musí být server, k němuž je úložiště připojeno, zapnutý a musí zde běžet zprostředkovávající služba. V případě souběhu více vstupně-výstupních požadavků a nedostatečného dimenzování výkonu procesoru dochází k poklesu vybavení požadavků za jednotku času a klientské aplikace musejí na DAS úložiště čekat. Stejný negativní jev můžeme pozorovat v případě, kdy aplikace na tomto serveru běžící zatíží procesor do té míry, že se nedostává výpočetního výkonu na vybavení požadavků na práci s datovým úložištěm. Tato omezení lze do značné míry ovlivnit výběrem vhodného hardwaru s dostatečným výkonem a propustností sběrnic, výběrem serverového operačního systému, jeho optimalizací a sladěním s aplikacemi.

Další limitující faktor souvisí do značné míry s fyzickým provedením úložišť tohoto typu, je jejich omezená škálovatelnost. Částečně lze eliminovat toto negativum provedením například s oddělenou částí pro datová média resp. pevné disky, ale i takto provedené DAS má své limity, jež jsou ve srovnání s dalšími typy datových skladů poměrně nízké.

Problematická také zůstává bezpečnost a správa zařízení typu DAS při nasazení ve větším počtu. V souhrnu také typicky zaostává toto řešení v parametru "Up time" tj. v maximálním časovém intervalu, kdy je zařízení bez přerušení schopno poskytovat definovanou službu.

4.2. Distribuované

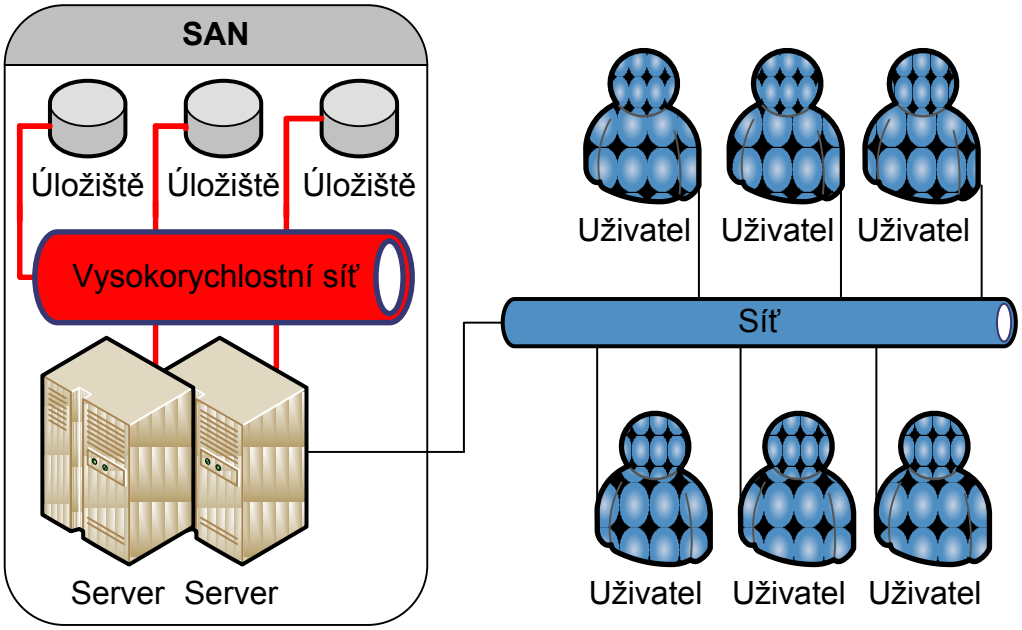
Do této kategorie spadají paměťové sklady založené na principu SAN – Storage Area Network a NAS – Network Attached Storage. Některé sofistikované přímo připojitelné DAS lze nakonfigurovat a připojit jako SAN nebo NAS a naopak lze některé zařízení určené pro práci v SAN nebo jako NAS zapojit třeba přímo k HBA serveru. Zařízení SAN a NAS mají společnou vlastnost a to, že jsou připojované ke klientům pře síťová rozhraní pomocí síťových

protokolů. Další jejich vlastnosti jsou pak natolik odlišné, že se každému budeme věnovat odděleně.

4.2.1. Storage Area Network (SAN)

Na rozdíl od DAS i NAS jsou datová úložiště v tomto případě připojená do dedikované vysokorychlostní sítě SAN optimalizované k tomuto účelu. Konsolidace diskového prostoru je navržena jako koncept centrálního a sdíleného diskového prostoru pro více serverů a překonává problémy, které s sebou přinášela technologie DAS. Zároveň je tato konsolidace odpovědí na otázku, jak řešit problém s nárůstem objemu ukládaných dat.

SAN je architektura připojující k serverům vzdálené paměťové sklady jevící se navenek jako lokálně připojené. Toto umožňuje existence sítě s vysokou propustností, oddělené od například LAN nebo WAN s připojenými klienty, která sama o sobě není souborově orientovaná, pracuje tedy s bloky dat. Na úrovni serverů mohou být datové bloky transformovány do souborů a dále pak takto distribuovány *viz Obr. 5*.



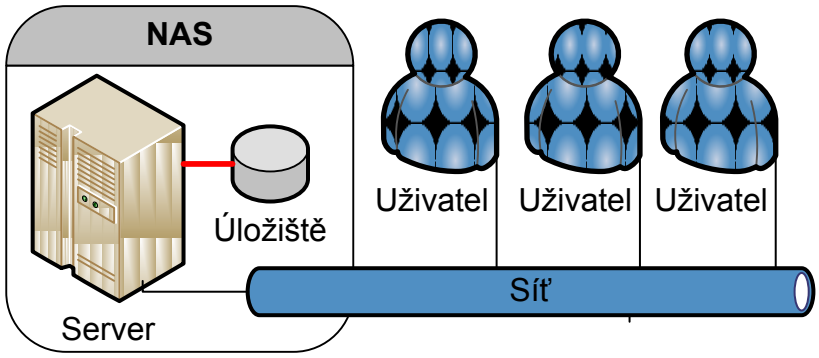
Obr. 5: Storage Area Network

Celý koncept je založen na jednoduché myšlence jednoho centrálního diskového úložiště, které je společné pro všechny uživatele a všechny typy dat. I když se tak navenek jeví, může být SAN velice rozsáhlé a komplexní. Toto úložiště díky své robustní architektuře, mechanismu zálohování dat a centrální správě výrazně zvyšuje nejen efektivitu práce s uloženými daty, zjednodušuje správu celého zařízení, ale zároveň pracuje i na mnohem

vyšší úrovni bezpečnosti. SAN mimo jiné umožňuje využití mechanizmů jako "disaster recovery" – obnova po selhání a implementaci vlastností jako například "snapshotting" – vytváření datového obrazu za chodu systému, nebo „volume cloning“ – vytvoření kopie svazku. Podrobněji budou některé z těchto metod popsány v kapitole 6.

4.2.2. NAS (Network Attached Storage)

Tvoří paměťový prostor připojený přes příslušné rozhraní do počítačové sítě *viz Obr. 6*. Klientům v heterogenní síti poskytuje přístup k souborům, nikoliv však k fyzickým blokům dat na paměťovém médiu.



Obr. 6: Network Attached Storage

Fyzicky tvoří NAS zařízení většinou počítač s nainstalovaným jednoúčelovým operačním systémem nebo jedním z běžných operačních systémů, nakonfigurovaným pro spolupráci s aplikací zajišťující funkce NAS. Z vnějšku se obě ovšem chovají stejným způsobem a to tak, že umožňují sdílení diskového prostoru systému. Teoreticky je možné, a v praxi jsou taková řešení vidět poměrně často, aby na těchto systémech byla spuštěna ještě jiná aplikace resp. služba podporující nebo doplňující služby NAS.

Konfigurace NAS zařízení probíhá většinou přes síťové rozhraní pomocí protokolu HTTP nebo TELNET a jejich klony. Z principu je možná konfigurace pomocí přímo připojené klávesnice a monitoru, ale výrobci dávají většinou přednost jednodušším uživatelským rozhraním, jako je třeba právě zmiňovaný internetový prohlížeč využívající protokol HTTP nebo jeho zabezpečenou verzi HTTPs.

5. Shrnutí a přehled typů datových úložišť

Z důvodu větší přehlednosti si uveďme důležité vlastnosti jednotlivých typů datových skladů zmiňované již dříve v textu ve zjednodušené formě v tabulce *Tab. 6*

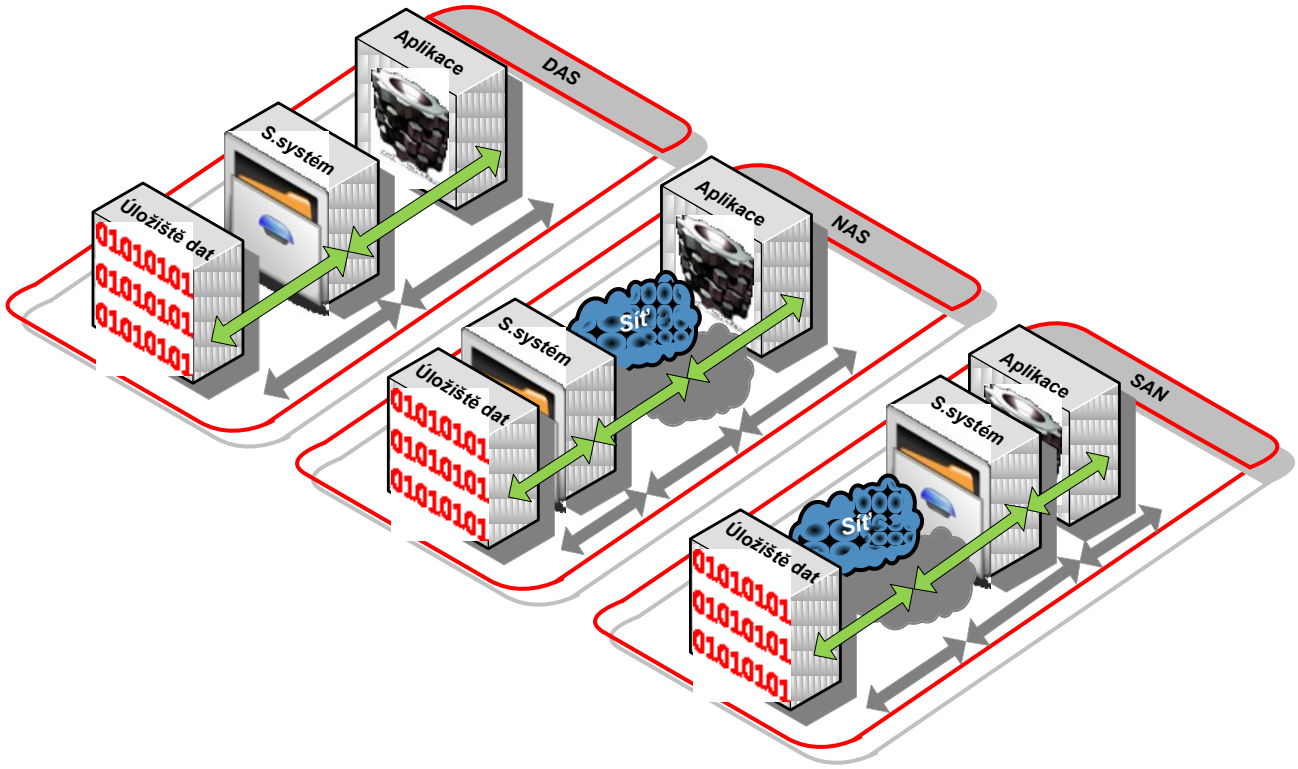
Z hlediska způsobu přístupu k datům můžeme definovat DAS zařízení jako samostatné zařízení, kde je datové úložiště spojeno s aplikačním serverem přímo, a to relativně krátkou sběrnicí. Aplikace má přitom přímý přístup k blokům dat nebo zprostředkovaně přes souborový systém.

V případě zařízení zapojeného jako NAS přímý přístup k datovým blokům není a data jsou aplikaci prezentována ve formě souborů, které jsou k aplikaci následně přenášena přes připojenou síť.

	SAN	***NAS***	***DAS***
Připojení do	*dedikované sítě s vysokou propustností*	*LAN (sdílená s klienty)*	*samostané zařízení*
Připojovací rozhraní	*FC, Ethernet popř. iSCSI apod.*	*Ethernet*	*ATA, SCSI, USB apod.*
Nosné médium	*Nejčastěji optický kabel, jinak vysokorychlostní metalické vedení*	*Většinou metalické vedení*	*metalické krátké vedení*
Použité protokoly	*Blokově orientované (ATA,SCSI)*	*Souborové (NFS,SMB)*	*ATA,SCSI*

Tab. 6: Shrnutí typů datových skladů

V posledním diskutovaném systému - SAN jsou přes síť pomocí vhodného protokolu přenášeny rovnou bloky dat, které jsou pak na opačném konci teprve umisťovány do souborového systému, kde jsou připraveny pro aplikační vrstvu. Vše řečené je znázorněno graficky v *Obr. 7.*



Obr. 7: DAS, NAS, SAN - Typy datových skladů

6. Základní úrovně RAID

V předešlých kapitolách byly diskutovány typy datových skladů z hlediska topologie a možného připojení vlastního paměťového prostoru, nyní bych se rád zaměřil na vlastní paměťové médium, jež je základem, každého takovéhoto úložiště.

Dále v textu zúžíme chápání pojmu paměťového media jako pevného disku. Pevný disk založený na mechanickém pohybu je v současné době používán v drtivé většině obecně použitých datových úložišť. Pro speciální účely mohou být pole mechanických disků nahrazeny disky SSD založené na statických pamětích FLASH. Tato pole poté vynikají řádově lepšími parametry, ale na druhou stranu jsou velice nákladná a jejich použití se musí zvažovat z ekonomicko-technického pohledu. FLASH pole bývají využity jako rychlá cache pro fragmentovaná a nejpoužívanější data. Pokud jsou data delší dobu nepoužita, přesouvají se automaticky do úložišť tvořených mechanickými pevnými disky a ty se ještě mohou po určité době zazálohovat resp. archivovat na paměť tvořenou třeba páskovými roboty apod.

RAID úrovně definované v dokumentu *lit. [1]* byly postupem času rozšířeny o další modernější architektury, proto hovoříme-li dnes o základních úrovních RAID, uvažujeme kromě zmiňovaných pěti také ještě dvě, a to úroveň RAID 6 a RAID 10. Tyto dvě úrovně jsou v současné době spolu s úrovní RAID 5 nejpoužívanější redundantní systémy, a proto je také zařadíme do podrobnějšího přehledu.

Dále si definujme některé pojmy, které budeme v dalším výkladu používat:

6.1. Odolnost proti chybám

Základní odolnost proti chybám ve světě úložišť znamená, že data zůstanou nedotčená, i když jeden z pevných disků selže. Systémy vyšší úrovně dovolují dokonce selhání dvou a více disků současně bez ztráty dat. Existují také enterprise – podnikové systémy, které dovolují selhání celé cesty, například řadiče samotných disků, konektorů, kabeláže dokonce výpadek napájení systému bez ztráty poskytované služby nebo dat. Redundance systémů nebo cest nespadá do RAID technologie, ale je považována také za formu odolnosti úložiště proti chybám.

6.2. Prokládání – Striping

Je takový algoritmus ukládání resp. rozprostření dat, který využívá dva a více disků, na které ukládá současně stejně velké bloky původních dat jakoby v proužcích – stripes. Tímto

způsobem se rychlost čtení a zápisu dramaticky zrychlí, jelikož systém zapisuje, resp. čte v jeden okamžik data ze dvou disků.

6.3. Zrcadlení – Mirroring

Je jednoduché kopírování dat z jednoho zdroje na dvě místa současně, resp. synchronizace dvou datových prostorů a vytvoření dvou identických a ve všech směrech zastupitelných kopií. Přičemž při zápisu se rychlost toku zapisovaných dat nijak nemění, ale při čtení již uložených dat je možné číst data paralelně ze všech kopií současně a tím řádově zvýšit datovou propustnost v závislosti na počtu zrcadel.

6.4. Parita

V úrovních RAID 2,3,4,5,6 a hybridních, kde jsou vyjmenované základní úrovně obsaženy je použita k obnově poškozených dat tzv. parita *viz (6.1)*. Parita je v našem případě redundantní informace vypočítaná na základě chráněných dat a pomocí níž lze předem určenou část chráněných dat zpětně dopočítat *viz (6.2)*. Vezměme si jako příklad RAID úrovně 3 a ukažme si, jakým způsobem probíhá rekonstrukce dat z paritních dat:

Uvažujme diskové pole sestávající z pěti disků: X0 – X3 obsahují datové bity , X4 – obsahuje paritní bit

Schéma tvorby paritního bitu je následující:

$$X4(i) = X3(i) \text{ xor } X2(i) \text{ xor } X1(i) \text{ xor } X0(i) \qquad (6.1)$$

Tímto jsme získali paritní bit – redundantní informaci, nyní předpokládejme, že disk X1 selhal, to znamená, jeho data nejsou nadále k dispozici. K oběma stranám rovnice přičteme

$$X4(i) \text{ xor } X1(i)$$

Dostáváme a dále upravíme:

$$X4(i) \text{ xor } X4(i) \text{ xor } X1(i) = X3(i) \text{ xor } X2(i) \text{ xor } X1(i) \text{ xor}$$

$$X0(i) \text{ xor } X4(i) \text{ xor } X1(i),$$

X4(i) xor X4(i) je vždy 0, stejně tak X1(i) xor X1(i)

Hodnotu bitu z disku, který má poruchu, získáme ze vztahu:

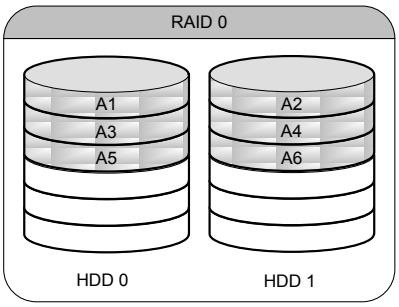
$$X1(i) = X4(i) \text{ xor } X3(i) \text{ xor } X2(i) \text{ xor } X0(i) \quad (6.2)$$

Výše popsaný princip výpočtu se používá pro RAID 3 až RAID 5 při výpadku jedno z datových disků pole, kdy pole přejde do takzvaného redukovaného režimu a chybějící data dopočítává za běhu. Takto získaná data se také zapisují na náhradní disk, který nahradí disk v poruše, aby se pole opět uvedlo do standardního režimu.

V úrovních 2 a 6 zůstává řešení stejné s tím rozdílem, že v RAID 2 se paritní informace počítá pomocí Hammingova kódu a pro úroveň číslo 6, která obsahuje parity dvě, se získává pomocí kódu Reed-Solomon.

6.5. RAID 0

Vlastně nenaplňuje význam slova „Redundant“ ze zkratky RAID, jelikož neposkytuje jakoukoliv redundanci a je tedy speciálním případem. Jediné, co poskytuje, je proklad, tj. rychlejší čtení resp. zápis. Příchozí data se ukládají na jednotlivé členy pole v jakýchsi proužcích (anglicky stripes) a tyto mohou být ukládány a čteny současně *viz Obr. 8,* což má velký vliv na zvýšení výkonu.

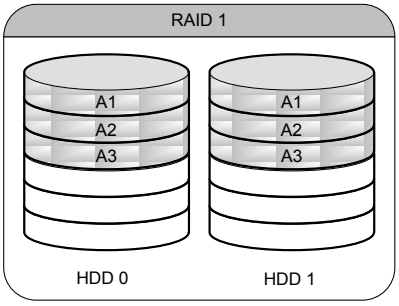


Obr. 8: Datová mapa RAID 0

6.6. RAID 1

Oproti předcházející konfiguraci poskytuje úroveň RAID 1 stoprocentní redundanci, jelikož vytváří datovou kopii disku 0 na disku 1 a obráceně – udržuje všechny disky bitově shodné *viz Obr. 9.* Co se rychlosti čtení a zápisu týče, řídí se parametry zápisu podle

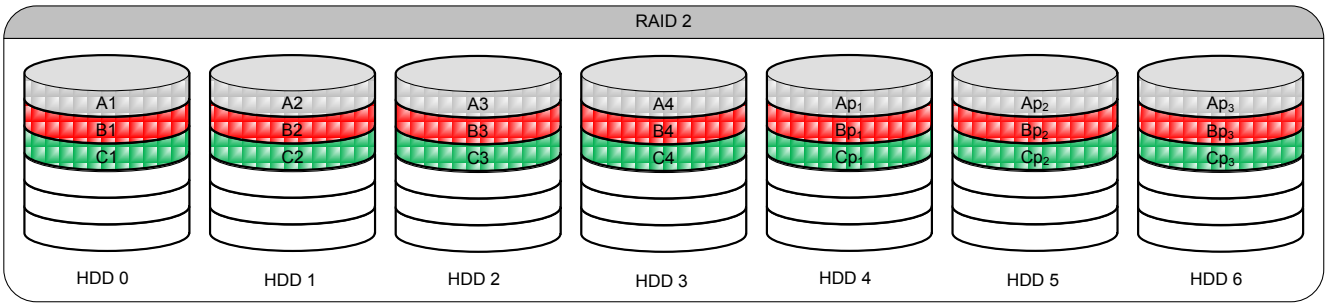
nejpomalejšího z disků v poli, takže k nárůstu výkonu nedochází, spíše zde můžeme pozorovat jeho mírnou degradaci. Tato penalizace je však vyvážena stoprocentní redundancí, takže po selhání jednoho ze členů pole zde máme k dispozici teoreticky bitově stejné další členy zmiňovaného pole a rychlostí čtení, jež je díky paralelismu násobná v závislosti na počtu disků v poli.



Obr. 9: Datová mapa RAID 1

6.7. RAID 2

Za této konfigurace přistupuje řadič k diskům bitově a zapisuje, resp. čte jednotlivé disky po bitech. Kombinací informací z jednotlivých disků posléze skládá slabiky a slova datových bloků. Počet použitých disků předurčuje použitá metoda zabezpečení dat proti chybám, jež je v tomto případě Hammingův kód. Pro Hamming(3,1) jsou tedy použity tři disky – dva paritní a jeden datový, Hamming(7,4) *viz Obr. 10.* používá sedm disků, z nichž čtyři jsou datové, zbytek je použit pro ukládání parity.



Obr. 10: Datová mapa RAID 2

Pozorný čtenář jistě odhadl, že tato úroveň nevyniká v nižších počtech disků příliš velkou úsporou datového prostoru a připočteme-li k tomu, že disky musí (pokud má být pole výkonné) hardwarově tuto úroveň podporovat, aby mohli synchronizovat úhlové rychlosti, nevyváží tuto skutečnost ani fakt, že v konfiguraci RAID 2 lze opravit selhání jednoho disku a to tzv. „On the fly“ (za letu).

Výkon citované úrovně není nikterak oslnivý a byl by vhodný, díky paralelismu, maximálně pro sekvenční čtení. Ostatní pracovní režimy nepodávají dostatečný výkon díky bitové orientaci, výpočtu parity a synchronizaci disků.

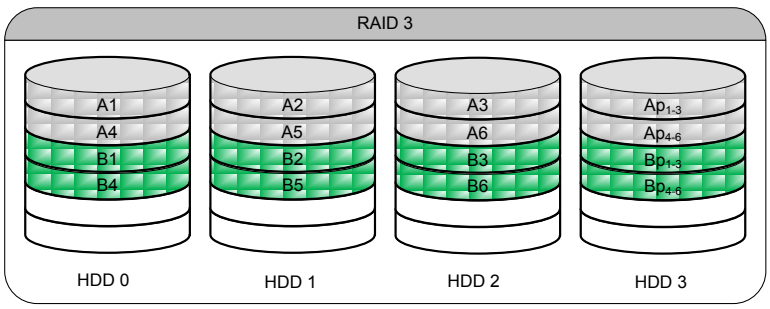
RAID 2 se díky nevalnému výkonu a vysokým nákladům na hardware ve větší míře nepoužívá.

6.8. RAID 3

RAID v úrovni 3 *viz Obr. 11* pracuje na shodném principu jako v úrovni 2, ale nepoužívá bitově orientovaný proklad, nýbrž bajtový.

Z RAID 2 má shodný vyhrazený paritový disk a stejně tak negativní vlastnost, která se s jeho použitím pojí, tj. úzké hrdlo v podobě právě paritového disku, na který je třeba zapsat a číst při všech vstupně výstupních požadavcích. Z tohoto důvodu, stejně tak jako výše uvedená úroveň, není vhodný pro vybavení mnoha nesourodých, krátkých, po sobě rychle následujících požadavků, tolik typických třeba pro provoz transakčních databází.

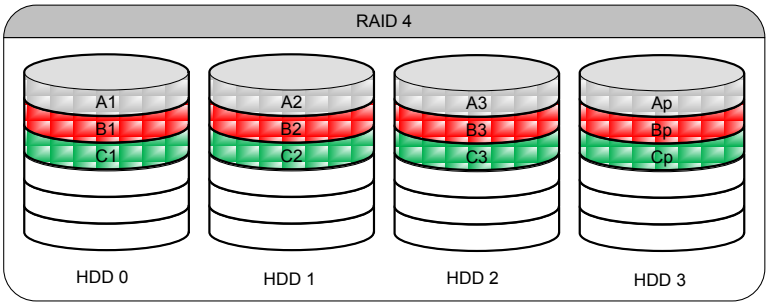
V současné době se tato úroveň RAID samostatně prakticky nepoužívá, setkat se s ní můžeme, i když ojediněle, v hybridních úrovních RAID.



Obr. 11: Datová mapa RAID 3

6.9. RAID 4

Úroveň je v mnoha směrech příbuzná úrovni 5 a 3. Na rozdíl od úrovně 3 je RAID 4 blokově orientovaný, což se projevilo na nárůstu výkonu při potřebě náhodného přístupu. Úroveň RAID 4 *viz Obr. 12* ovšem stále používá vyhrazené disky pro uchovávání paritních dat a nedistribuuje je na všechny členy pole jako je tomu u RAID 5. Tento fakt ovšem předurčuje úroveň 4 být pomalejší, jelikož na paritním disku vzniká „úzké hrdlo“ a při masivním počtu náhodných vstupně-výstupních požadavků je nutné čekat na vybavení požadavků disku s paritními daty.

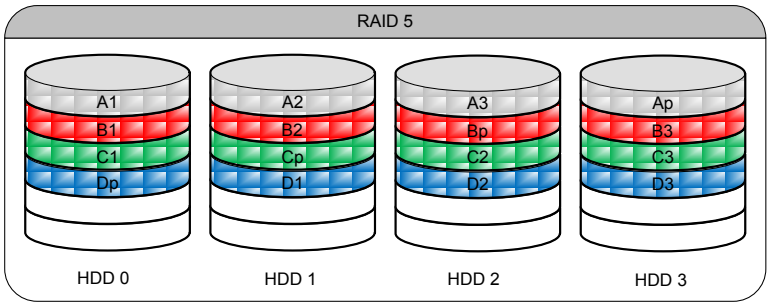


Obr. 12: Datová mapa RAID 4

6.10. RAID 5

RAID úrovně 5 *viz Obr. 13* je jeden z nejpoužívanějších úrovní současnosti. Poskytuje dobrý výkon pro databázové aplikace v transakčním režimu, ale je schopno pracovat i v sekvenčním provozu a dodávat data velkou rychlostí. Svou oblíbenost získalo především dobrým poměrem redundance k výkonu, protože zatímco třeba RAID 1 je schopno poskytnout velikost pole jako nejmenší disk v tomto poli, RAID 5 poskytuje kapacitu o jeden disk menší krát kapacitu nejmenšího disku. Pro představu v nejmenší konfiguraci tří stejných disků je kapacita pole dvě třetiny kapacity disku.

Není bez zajímavosti, že RAID 5 se také někdy používá bez redundance jako pomalejší RAID úrovně 0 avšak s možností rozšířit kapacitu dle aktuálních požadavků.



Obr. 13: Datová mapa RAID 5

RAID 5 (a nejen ono) má ovšem i několik nectností, pro které je částí uživatelů *lit. [11]* zatracováno. RAID v úrovni 5 není například odolný vůči chybě zvané „write hole“, kdy např. z důvodu selhání napájení, nebo chybě disku nejsou zapsány korektní data nebo parita. Pole tuto chybu nedetekuje a v případě selhání disku jsou z chybné parity rekonstruovány chybná data. Částečně řeší tuto chybu vyrovnávací-cache paměť zálohovaná baterií, ale zde dochází opět k negativním syndromům při extrémním počtu vstupně výstupních operací, jež pole není schopné vybavit a cache paměť se zaplní, dochází v tomto okamžiku k rapidnímu poklesu výkonu, jelikož plná paměť neplní svojí funkci. Zálohované cache paměti bývají velice nákladnou součástí kontrolerů a prodražují tak celkové řešení. Jinak obchází „write hole“ syndrom Z-RAID architektura. Ta využívá implementaci redundance na bázi souborového systému ZFS a pro malé zápisy bloky nemodifikuje, nýbrž vytvoří kopii bloku a poté kopii pouze označí za validní a stará data vymaže.

Další negativní vlastností je chování RAID 5 v degradovaném stavu, tj. po selhání jedné jednotky, kdy při každém zápisu je nutné zapsat na všechny disky. Připočteme-li k tomu ještě např. čtení pro rekonstrukci dat na záložní disk, stává se výkon systému přinejmenším diskutabilním.

Jak jsme již uvedli, obliba RAID 5 dosáhla svých rozměrů především kvůli nákladům na jednotku diskové kapacity. To již přestává být motivací, jelikož jednotka diskové kapacity zlevnila natolik, že lze uvažovat o RAID úrovních s větší spotřebou fyzického paměťového prostoru na totožný virtuální disk jako pro RAID 5.

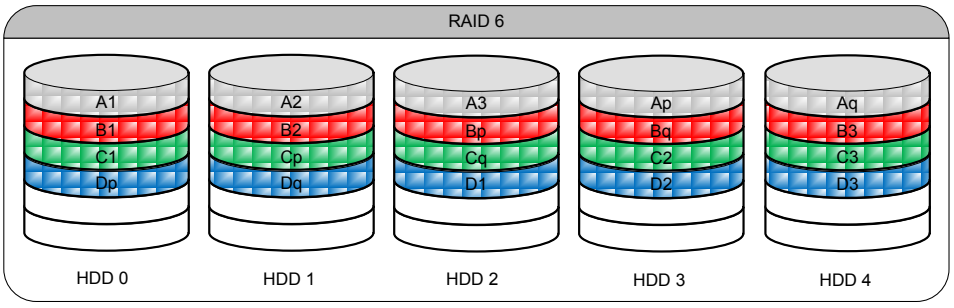
6.11. RAID 6

Pokud řekneme, že některé úrovně RAID jsou si až na detaily podobné, platí toto právě o úrovni 6, která je totožná co do architektury a hlavní myšlenky s architekturou úrovně 5. RAID 6 rozšiřuje RAID 5 pouze o další redundanci *viz Obr. 14*, která zabezpečí systém proti

výpadku i druhého disku. V důsledku to znamená, že architektura zdědila po předchůdci všechny své vlastnosti, jak kladné tak záporné. Úroveň 6 tedy zůstává blokově orientovaná, prokládaná přes všechny disky pole ovšem s duální paritou. Parita je taktéž rozprostřena přes všechny disky pole, ovšem zabírá, díky své dualitě, kapacitu dvou disků zmiňovaného pole.

Změnou v této úrovni je také algoritmus výpočtu parity. Parita P+Q se v tomto případě počítá pomocí Reed Solomon algoritmu.

Potřeba odolnosti systémů při ztrátě dvou disků narůstala s velikostí zpracovávaných dat a počtem disků v poli. Disky v konkrétním poli bývají osazovány z důvodu přibližně stejných parametrů od jednoho výrobce a tyto vyrovnané parametry se pak podílejí na lepším výkonu celého pole. Ovšem zmiňované disky mají také stejný parametr MTTF a co více většinou není rozptyl TTF – doby do poruchy stejné série disků veliký tak, jak bychom dnes potřebovali. Při výpadku jednoho z disků i při použití RAID 5 a SPARE – záložního disku trvá při současných kapacitách a propustnosti disků rekonstrukce pole i třeba půl dne, což je doba, po kterou není systém s RAID 5 nikterak chráněn a výpadek druhého disku by měl fatální následky. Často se také stává, i když je to s podivem, že pokud není osazen SPARE disk, musí operátor disky vyměnit a chybnou identifikací špatného disku vymění obsluha „zdraví“ disk. Následky a důvod vzniku duální parity jsou zřejmé.



Obr. 14: Datová mapa RAID 6

RAID 6 bývá implementován do high-end systémů s vysokou dostupností, ale v méně robustních podobách jej najdeme i v SOHO zařízeních pro použití v domácích a malých kancelářích. Dnes ovšem stále častěji zaznívají názory *lit. [12]*, že do budoucna bude, z důvodu velkých kapacit polí, nutné s dual parity přejít na paritu triple, tj. zvětšit redundanci polí o ještě jeden disk.

6.12. Shrnutí vlastností základních úrovní

Sumarizaci nejdůležitějších vlastností poskytuje tabulka *Tab. 7.*

Úroveň	Popis	Min.počet HDD	Využitelnost prostoru	Odolnost proti chybám	Spolehlivost	Rychlost čtení	Rychlost zápis
RAID 0	Proklad/rozložení dat na jednotlivé disky	2	1	0 (žádná)	1−(1−r)n	nX	nX
RAID 1	Zrcadlení bloků dat	2	1/n	n−1 disk	rn	nX	1X
RAID 2	Proklad/rozložení bitů na jednotlivé disky s paritním diskem(Hamming)	3	1 − 1/n · log2(n-1)	Detekuje a opraví chybu jednoho disku	závislá na konfiguraci	závislá na konfiguraci	závislá na konfiguraci
RAID 3	Proklad/rozložení bajtů na jednotlivé disky s paritním diskem	3	1 − 1/n	1 disk	n(n−1)r2	(n−1)X	(n−1)X*
RAID 4	Proklad/rozložení bloků na jednotlivé disky s paritním diskem	3	1 − 1/n	1 disk	n(n−1)r2	(n−1)X	(n−1)X*
RAID 5	Proklad/rozložení bloků dat i paritových dat na všechny disky	3	1 − 1/n	1 disk	n(n−1)r2	(n−1)X*	(n−1)X*
RAID 6	Proklad/rozložení bloků dat i paritových duálních dat na všechny disk	4	1 − 2/n	2 disky	n(n-1)(n-2)r3	(n−2)X*	(n−2)X*
RAID 10	Proklad/rozložení bloků dat na vnořené zrcadlené disky	4	1/n	1disk ze zrcadlené skupiny	1−(1−rn)n	nX*	nX*

Tab. 7: Shrnutí vlastností základních úrovní lit. [13]

n – počet datových disků v poli
r – chybovost (%/100)
X – datová propustnost jednoho disku
** - hardware musí podporovat*

7. Vnořené (hybridní) úrovně RAID

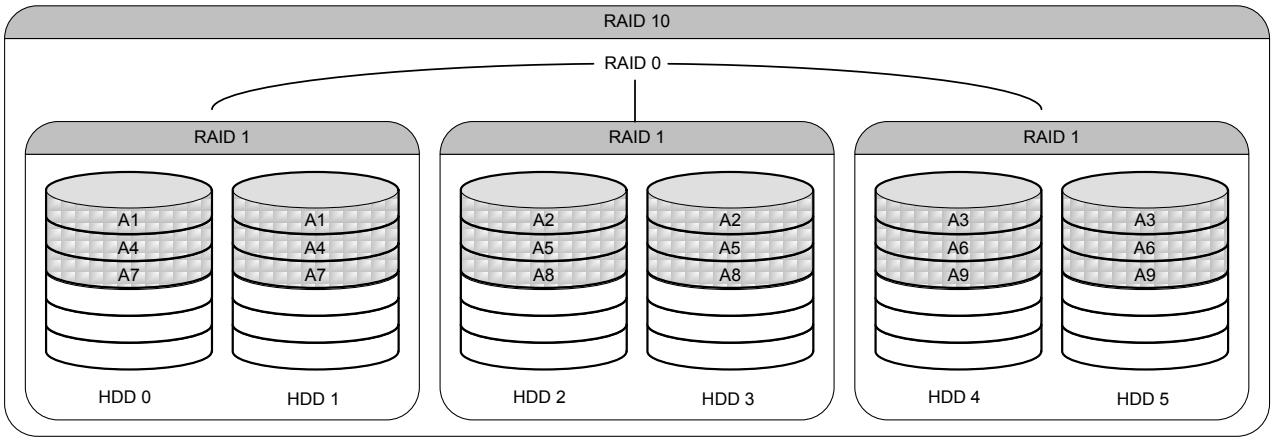
Kromě základních úrovní architektury RAID existují úrovně vnořené, tzv. hybridní. Tyto využívají vlastností jednotlivých základních úrovní a vhodně je mezi sebou kombinují tak, aby bylo dosaženo pole požadovaných parametrů.

Hybridní úrovně mohou obsahovat dvě i více úrovní, ale více nežli tři vnořené základní úrovně se nepoužívají. Kombinací je nepřeberné množství, my se proto soustředíme jen na ty smysluplné a používané.

Názvosloví pro hybridní úrovně se ustálilo na formátu RAID [číslo úrovně nejvíce vnořené]...[číslo úrovně nejvýše postavené], nebo skoro totožné RAID [číslo úrovně nejvíce vnořené]+...+[číslo úrovně nejvýše postavené]. Příklad: RAID 100 nebo RAID 1+0+0 reprezentuje několik polí úrovně 1 sloučených do polí úrovně 0 a tyto pak konečně sdruženy do výsledného pole RAID 0.

7.1. RAID 10

Několik polí úrovně 1 sdružených pomocí úrovně 0 vytváří pole RAID 10

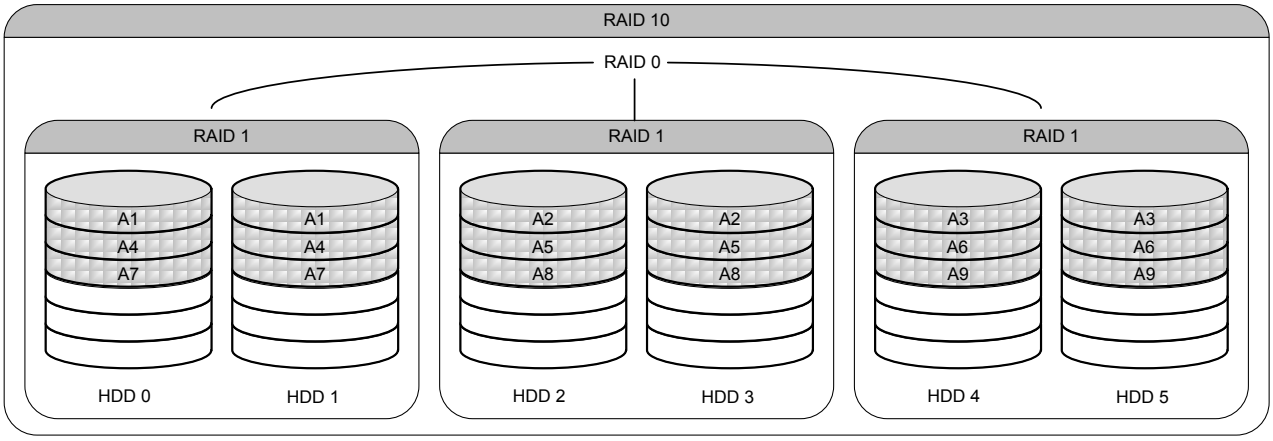


Obr. *15*. Díky klesající ceně za diskový prostor a výborným výkonnostním parametrům se tato úroveň začíná více prosazovat a je důstojnou konkurencí RAID 5. Ve výkonnostních parametrech poráží všechny doposud zmiňované RAID úrovně, vyjma RAID 0. Hodí se jak pro sekvenční čtení a zápis, tak pro aplikace s velkým množstvím požadavků na změny.

Mluvíme-li o výkonu, musíme připomenout, že některé její implementace umožňují dále konfigurovat a ladit už tak výborný výkon. Děje se tak prostřednictvím změny pořadí ukládaných bloků dat na fyzické pevné disky, jež určuje datový systém DDF (Disk Data Format) používaný na těchto polích a fyzicky umístěný u středu disku. Změnou pořadí, resp. pozice datových bloků na jednotlivých fyzických discích v poli úrovně 1 dokážeme

optimalizovat pohyb čtecích hlav tak, že například: Budou-li zrcadlené kopie datového bloku umístěné na disku č.1 na vnějšku média a na disku č.2 na vnitřní stopě, bude se při čtení tohoto bloku rozhodovat, zda je aktuálně hlava blíže vnitřní nebo vnější stopě a podle toho se zvolí disk, ze kterého se bude daný blok číst.

Odolnost systému proti selhání určuje počet zrcadlených disků ve skupině. Při použití dvou zrcadlených disků je výsledné pole odolné pouze proti selhání jednoho disku ve skupině a po selhání i druhého disku dochází ke ztrátě všech dat v poli.



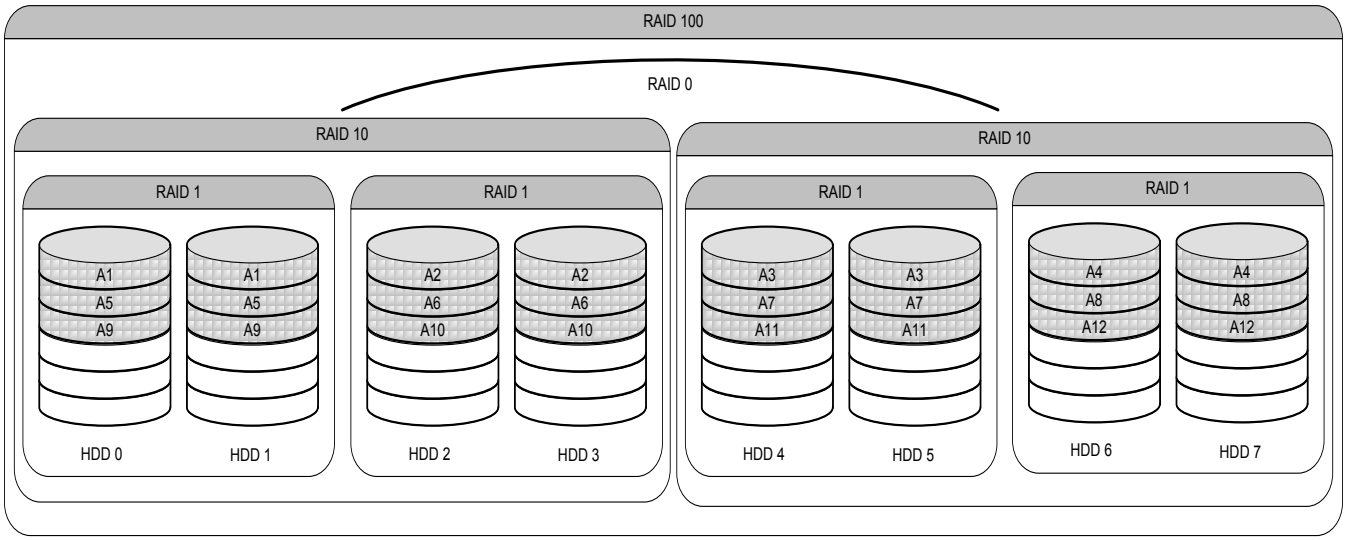
Obr. 15: Datová mapa RAID 10

Sesterská úroveň RAID 01, která má prohozené úrovně (tj. zrcadlí se prokládaná pole), má velice podobné parametry s jedním rozdílem. Při selhání disku a jeho následné výměně je u této úrovně nutné více aktivity. Nezrcadlí se pouze ve skupině, jako je tomu u RAID 10, nýbrž se zrcadlí celé prokládané skupiny – je tedy nutné přenést větší objem dat. I když se přenáší větší rychlostí, je tato operace časově náročnější než u RAID 10.

7.2. RAID 100

Označuje stripovaná pole úrovně 10. Vyniká vysokým výkonem stejně jako RAID 10, tudíž se hodí pro aplikace vyžadující jak velkou propustnost velkých bloků dat, tak pro aplikace databázové a multiuživatelské.

Používá se velice rozsáhlých enterprise polích, kde kapacitu limituje počet kanálů RAID 10. Tyto pole pak lze opět zařadit do pole úrovně 0 a tím dostáváme úroveň RAID 100 s dramaticky větší kapacitou. V minimální konfiguraci obsahuje pole osm disků *viz Obr. 16.*

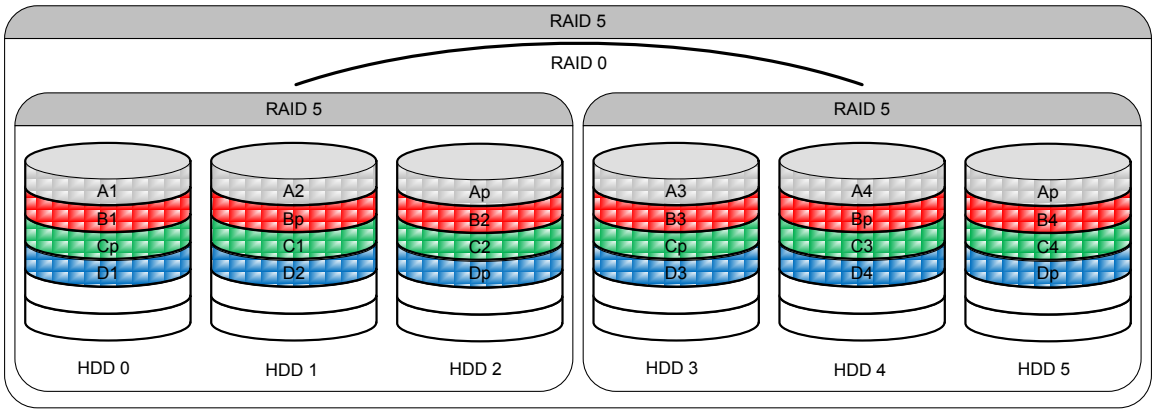


Obr. 16: Datová mapa RAID 100

7.3. RAID 50

Sestává ze stripovaných polí RAID 5 *viz Obr. 17*. Zvyšuje výkon pole základní úrovně, zejména pak při zápisu, kde je rozdíl markantní. Je vhodný pro aplikace s náhodným přístupem k datům. RAID 50 neplýtvá příliš diskovým prostorem jako například RAID 10 a přitom poskytuje stejnou redundanci.

Slabinou tohoto hybridního pole je odolnost proti selhání, kdy při selhání jednoho disku ve skupině RAID 5 o třech discích je po dobu obnovení pole bez redundance a selhání dalšího disku ve skupině vede ke ztrátě dat. Čas v degradovaném režimu bývá většinou pro tato pole v řádu desítek hodin, což je poměrně velké riziko.



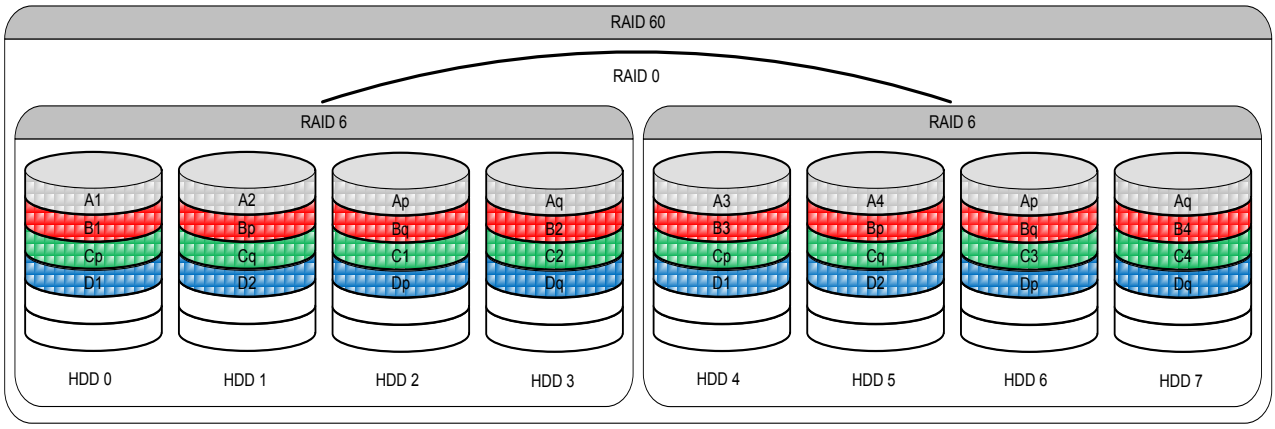
Obr. 17: Datová mapa RAID 50

7.4. RAID 60

Kombinuje prostý proklad a distribuovanou duální paritu RAID 6. Pole úrovně RAID 60 disponuje zlepšenou odolností proti selhání, jelikož například v minimální konfiguraci *viz Obr. 18* odolá selhání až poloviny disků, přičemž musí být vždy maximálně dva z jednoho subpole úrovně 6.

Striping pomáhá zvětšovat kapacitu a výkon bez zásahu do základních polí, kde by přidání dalších disků mělo za následek snížení výkonu jako důsledku čtení z celého pole při výpadku jednoho nebo dvou disků.

RAID 60 je v době vzniku tohoto dokumentu nejbezpečnější pole, schopné začlenit stovky disků při současném výborném výkonu i pro transakční, multiuživatelské aplikace a pro všechny aplikace, kde je nutné přistupovat často k malým blokům dat.



Obr. 18: Datová mapa RAID 60

V minimální konfiguraci potřebuje pole této úrovně osm fyzických disků a redundantní informace zabere čtyři disky.

7.5. Shrutí vlastností hybridních úrovní

Tabulka obsahuje nejdůležitější vlastnosti vnořených úrovní a jejich popis.

Úroveň	Popis	Min. počet disků	Využitelnost prostoru	Odolnost proti selhání	Rychlost čtení	Rychlost zápis
01	úroveň subpole RAID0, subpole sdružena do RAID1	3	1/G	G − 1 až m(G − 1)	(G · m)X	mX
10	úroveň subpole RAID1, subpole sdružena do RAID0	4	1/m	m − 1 až G(m − 1)	(G · m)X	GX
50	úroveň subpole RAID5, subpole sdružena do RAID0	6	(m − 1)/m	1 až G	G(m − 1)X*	G(m − 1)X*
60	úroveň subpole RAID6, subpole sdružena do RAID0	8	(m − 2)/m	2 až 2G	G(m − 2)X*	G(m − 2)X*
100	úr. subpole RAID1, subpole sdružena RAID0 a opět RAID0	8	1/m	m − 1 až G(m − 1)	(G · m)X	GX

Tab. 8: Shrnutí vlastností hybridních úrovní lit. [14]

G - počet subpolí
m –počet disků ve vnořeném subpoli
** - předpokládá se hardware s dostatečnou datovou průchodností*

8. Nestandardní a proprietární úrovně

8.1. IMST (Intel Matrix Storage Technology)

Je jedním z nejrozšířenějších, ačkoli ne nejpoužívanějších proprietárních architektur storage systémů. Rozpor je způsoben tím, že pomalu každý, kdo vlastní současnou stanici s procesorem společnosti Intel®, disponuje touto technologií, ačkoli jí v drtivé většině nevyužívá vůbec nebo ne naplno. Architektura *lit. [15]* je založená na standardních algoritmech základních úrovní RAID 0,1,5 a hybridní úrovni 10, přičemž umožňuje jejich současné použití na stejných fyzických discích. Stanice se dvěma harddisky, vybavená touto technologií tedy může mít uložená důležitá data jako například dokumenty na poli RAID 1, zatímco například swapovací soubor a nainstalované hry mohou být uloženy také na těchto discích, ale v datovém prostoru zařazeném do pole RAID 0.

8.2. RAID 1E

Data se ukládají vždy na dva sousedící disky současně, tím vzniká redundance, která je při použití již tří disků menší než je tomu u RAID 1, ovšem za cenu snížení odolnosti proti selhání na pouze jeden disk v poli. Zápis tedy probíhá srovnatelně rychle jako na samostatný fyzický disk, ovšem čtení probíhá s dvounásobným přenosem.

8.3. RAID 5E

Jedná se o modifikovanou úroveň pole RAID 5, která je rozšířena o nikoliv SPARE disk, nýbrž SPARE diskový prostor na konci každého adresního prostoru disku. Tento prostor je při selhání disku okamžitě k dispozici.

8.4. RAID 5EE

Dále modifikuje úroveň 5E, potažmo úroveň 5 změnou fyzického rozložení SPARE STRIP-ů a to tím způsobem, že s nimi zachází stejně jako s datovými. Rozdistribuovává je po všech discích v poli a tím snižuje čas pro inicializaci pole pod čas potřebný pro inicializaci pole úrovně 5E.

Distribuování SPARE prostoru zamezuje sdílení SPARE kapacity s ostatními poli v systému. Dále pak neumožňuje dva logické disky na jednom poli a zdaleka ne všechny řadiče tuto úroveň podporují.

8.5. RAID 6E

Je analogií k RAID 5E s duální paritou tzn., že využívá tazvaný „hot spare“ – rezervní SPARE diskový prostor na konci každého disku pole.

8.6. RAID 7

Ryze proprietární patentované řešení společnosti Storage Computer Corporation, které bylo dále přeprodáváno několika dalším společnostem vychází z principů RAID 3 a RAID 4. Výkonným hardwarem potlačuje limity polí s prokladem a dedikovanými paritními disky, aplikuje velké vyrovnávací paměti na několika úrovních. Velké nezálohované vyrovnávací paměti vyžadují stálé napájení a z tohoto důvodu je nutná aplikace nepřerušitelného zdroje napájení (UPS).

RAID 7 nenabízí z dnešního pohledu perspektivní řešení, ani technicky ani obchodně nenabízí velké možnosti dalšího rozvoje.

8.7. RAID Z

Zde zastupuje řadu redundantních systémů postavených výše nežli je fyzická vrstva tj. v našem případě pevné disky a zároveň níže než je operační systém. Jádro architektury je postavené na systému souborů ZFS – Zettabyte File System, který sám o sobě klade důraz na integritu uchovávaných a přenášených dat. Jeho vlastnosti budou více rozebrány v následující kapitole.

9. Nastupující technologie a predikce

9.1. SSD – Solid State Disk

Aktuálně nejdynamičtějším vývojem procházejí disky založené na principu nevolatil

ních pamětí, takzvané SSD disky. V současné době jsou jejich vlastnosti mírně přeceňovány, jelikož SSD disky trpí odlišnými, ale minimálně stejně závažnými, neduhy jako jejich technologicky starší soukmenovci. SSD disky například stárnou mnohem rychleji a jejich výkon se časem zhoršuje, jejich řadiče se musí neustále starat o mapování dat na disku tak aby nebyla určitá část paměťových buněk namáhána nadměrně. Také propustnosti se musí pomáhat, jelikož nativně je paměťová buňka SSD relativně pomalá, a to většinou implementací prokladu podobného RAID 0 u standardních disků. Řadič SSD se také musí starat o nepoužitá data na disku a musí vykonávat mnoho dalších funkcí. Z citovaného vyplývá, že řadič je stěžejní komponent a hlavně on rozhoduje o vlastnostech, jako jsou výkon, propustnost, počet vstupně výstupních operací za jednotku času, spolehlivost a doba, po kterou bude schopen garantovat tyto parametry.

Jelikož ani SSD disky nejsou bezporuchové, je třeba také vypíchnout rozdíl mezi nefunkčním HDD a SSD. Zatímco ze standardního disku lze ve většině případů, při vynaloženém úsilí úměrné poškození HDD, data zachránit, pokud nebudeme štěstí a u SSD neselhal „jen“ řadič je téměř jisté, že nebudeme moci zachránit tolik dat co z mechanického disku.

9.2. ZFS – Zettabyte File System

Reaguje na situaci, kdy podle výzkumu organizace CERN *lit. [16]* minimálně zhruba každý 10^{16}bit trpí syndromem „silent corruption“ tedy nedetekovanou tudíž neopravitelnou chybou. Fakt sám o sobě by nebyl nikterak dramatický, ovšem vezmeme-li v úvahu, že dnešní pole dosahují třeba i 120PB vyjde nám, že toto pole obsahuje přinejmenším jednu chybu. V rozsáhlých úložištích organizací poskytující cloudové služby jsou uživatelé schopní generovat takový datový provoz, že nedetekovaná chyba postihuje tyto systémy v intervalu řádů desítek minut a další příklady by mohli následovat.

ZSF chrání data počítáním parity, která zohledňuje i pořadí bloků, dále počítá hash (SHA-256), který ukládá mimo chráněný blok a při čtení jej počítá opět a porovnává s uloženým, pokud projde kontrolou předává datový blok dále, v opačném případě vyhledá redundantní data a pomocí logiky známé již z RAID architektury poškozená data obnoví. ZFS

může obsahovat duální redundanci nebo trojnásobnou redundanci. Při implementaci ZFS je doporučeno vyřadit všechny RAID hardwarové řadiče a umožnit ZFS přímý přístup na disk, nebo konfigurovat pole alespoň jako JBOD.

RAID Z, RAID Z2 RAID Z3 jsou jakousi nadstavbou ZFS a umožňují správu virtuálních zařízení, jejich spojování do větších celků a implementaci mirroring a striping technik. RAID Z je analogií k RAID 5, zatímco RAID Z2 je analogický s RAID 6 tj. obsahuje duální paritu pro selhání dvou disků. V roce 2009 byl RAID Z rozšířen o klon RAID Z3, který reaguje na potřeby další redundance při selhání dvou disků a zavádí trojnásobnou paritu.

9.3. Predikce vývoje

Z již uvedeného lze odvodit, že problémy způsobené nedetekovanými chybami budou narůstat spolu s nárůstem objemu zpracovávaných dat. Použití systému, který by byl schopen, když ne eliminovat, tak alespoň snížit jejich výskyt na minimum se zdá být nevyhnutelné. Současně použití levných RAID systémů s jinými úrovněmi než 1 vede téměř vždy ke ztrátě dat a je otázkou pouze času, kdy se tak stane.

Z těchto důvodů se zdá být ZFS a na něm postavené RAID Z dostatečně bezpečné a bude záležet pouze na konfiguraci, v jaké bude provozováno. Právě škálovatelnost je velkou výhodou, jelikož pole založená na této architektuře lze konfigurovat ve velkém rozsahu od prostého hashovaného úložiště až po téměř v běžné praxi nepoužitelné nastavení se zapnutou historií, třemi zrcadly, třemi paritami atd. atd.

ZFS plně podporuje SSD disky a právě tato kombinace bude hrát velkou roli v nově nasazovaných hybridních systémech, kde se uplatní vlastnosti jak standardních HDD, tak rychlých SSD. Tyto systémy by se měli rozšířit do centralizovaných úložišť jako součást virtualizovaných cloudových data center. Uživatel cloudové služby dostane k dispozici službu, jejíž součástí bude i datový prostor, který bude vyhrazen ve virtuálních diskových zařízeních, jejichž fyzické umístění a topologie zůstane před uživatelem skryta. Zde se budou data ukládat pod dohledem ZFS na SSD a zrcadlit do geograficky odděleného místa.

ZFS bylo koncipováno s velkým výhledem do budoucnosti, na což ukazuje například fakt, že je schopno obsloužit 256 kvadrilionů ZB datového prostoru. Jinak je tomu u SSD disků, ty narážejí již dnes na své limity, které se musí různými způsoby obcházet, z uvedeného se dá spekulovat, že technologie paměťové buňky použitá v SSD by měla být postupně nahrazena technologií výrazně vhodnější, ať již na podobném principu nebo např. na principu memristoru nebo jiné, prozatím nerozšířené technologie.

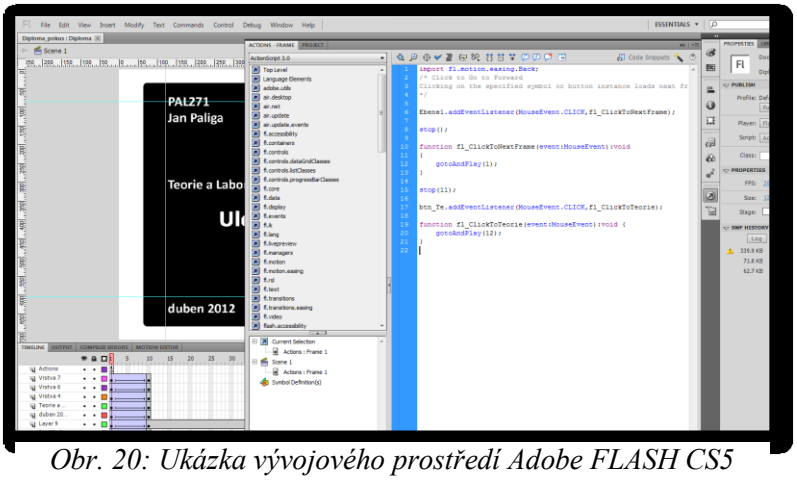
10. FLASH prezentace

Jako pomůcka při studiu tématu RAID polí slouží interaktivní FLASH aplikace *viz Obr. 19* zabývající se touto tématikou. Aplikace je interaktivním vodítkem s vysvětlením základních pojmů a principů RAID architektury a zároveň průvodcem jednoduchou laboratorní úlohou.

Obr. 19: Ukázka grafického zpracování FLASH aplikace

10.1. Vývojové prostředí

Pro vývoj aplikace byly zvoleny nástroje z dílen společnosti Adobe a Microsoft. Jedná se především o produkty:



Obr. 20: Ukázka vývojového prostředí Adobe FLASH CS5

Adobe Flash CS5 CZ – obsahuje hlavní vývojové prostředí *viz Obr. 20* a integruje programovací jazyk Action Script 3.0 do grafického návrhu. Poskytuje široké možnosti exportu jak do formátů, jež lze snadno implementovat do HTML stránek a zpřístupnit pomocí www služby, tak do samostatně spustitelných souborů.

Adobe Photoshop CS5 CZ – byl použit pro tvorbu grafických prvků aplikace spolu s

Microsoft Visio 2007 – ve kterém byly vytvářeny použité schémata a diagramy.

Nativně je v prostředí FLASH CS5 používán programovací jazyk Action Script 3.0, což je plně objektově orientovaný jazyk asi nejvíce podobný známému jazyku Java.

10.2. Popis aplikace

Po spuštění aplikace se zobrazí úvodní stránka s informacemi o díle. Aktivací tlačítka SPUSTIT přejdeme na stránku výběru aplikace. Na této stránce se zobrazí standardní navigační lišta obsahující tlačítka ZPĚT OBSAH a DÁLE. Navigační lištu máme k dispozici v průběhu celé aplikace. Volbou aplikace se dostáváme na obsah, který můžeme přeskočit pomocí navigačního tlačítka nebo zvolit požadovanou kapitolu ze seznamu. Prostřednictvím navigační lišty procházíme oběma aplikacemi až do konce.

11. Návrh laboratorní úlohy

Cílem laboratorní úlohy s názvem Ukládání dat v polích RAID je seznámit studenty s existencí, druhy a základními vlastnostmi polí RAID. Dále budou studenti provedeny praktickou ukázkou ukládání dat do zařízení NAS vybaveného polem RAID.

Při provádění laboratorní úlohy se předpokládá elementární znalost TCP/IP sítí, ethernetové topologie a práce s operačním systémem Windows.

Sestavení jednotlivých úrovní RAID polí je až na výjimky časově náročnou procedurou, která vyžaduje časy řádově kolem sedmi až osmi hodin. Potřebný čas stoupá s kapacitou použitých disků resp. disku s nejmenší kapacitou v poli. Rozsáhlým měřením bylo ověřeno, že ani s nejmenším diskem, jež je mimochodem pro SATA provedení cca 120GB, nebude dosaženo časů, které by byly akceptovatelné pro laboratorní úlohu v rozsahu jedné vyučovací hodiny *viz Tab. 9.*

RAID	HDD	500GB	250GB
0	1,2,3,4,5	15min	
1	1,2,3,4,5	420min	
1	1,2	170min	80min
5	1,2,3,4,5	540min	
5	1,2,3	350min	
6	2,3,4,5	500min	
10	1,2,3,4	500min	

Tab. 9: Thecus N5200 - Naměřené časy inicializace pole pro některé konfigurace

Ve světle citovaného faktu vyvstává potřeba nesestavovat jednotlivá pole v hodině, nýbrž je mít již sestavená. Tohoto dosáhneme buď výměnou sadou disků, na kterých je již konkrétní úroveň pole vytvořena, nebo bez výměny, kdy pole vyžadující dlouhý sestavovací čas

bude již sestavené z minulé vyučovací hodiny. Na základě zmiňovaného bude vytvořen variantní návrh.

11.1. Varianta A – S výměnnými sadami HDD

Pole JBOD a RAID úrovně 0 lze vytvářet v průběhu hodiny, RAID v úrovních 1,5,6 a 10 je nutné předpřipravit na sady disků. Počtu úrovní odpovídá počet potřebných sad tj. pro prezentaci všech úrovní je zapotřebí tří sad disků:

- Sada pro RAID úrovně 6 obsahuje všech pět disků (4 disky – min.počet disků pro tuto úroveň + 1 SPARE disk).
- Sada pro RAID úrovně 5 obsahuje čtyři disky (3 disky – min. počet disků pro tuto úroveň + 1 SPARE disk)
- Sada pro RAID úrovně 1 obsahuje tři disky (2 disky – min. počet disků pro tuto úroveň + 1 SPARE disk)
- Sada pro RAID úrovně 10 obsahuje také všech pět disků (4 disky – min. počet disků pro tuto úroveň + 1 SPARE disk)

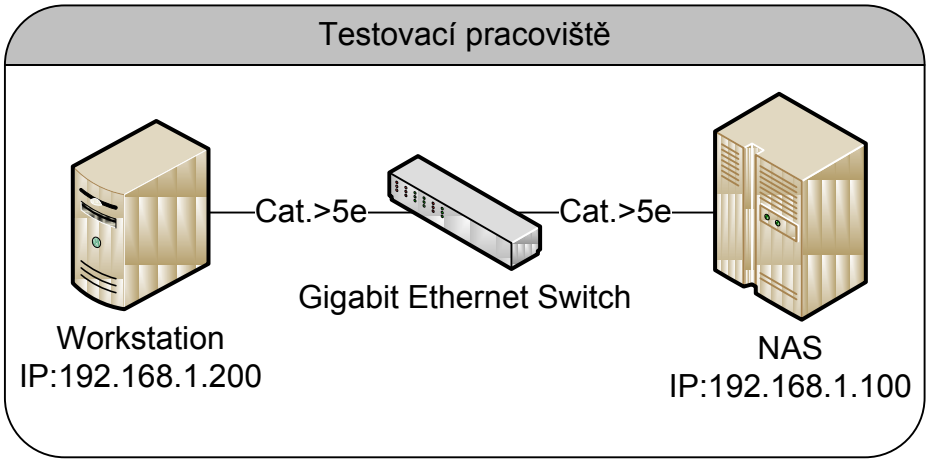
Celkem je tedy potřeba 12kusů disků nejlépe stejné kapacity a šarže. Kapacita disků je libovolná a může být i minimální možná (cca. 120GB). Náklady na zakoupení disků lze zmenšit vynecháním SPARE disků u všech nebo jen u některých úrovní až o čtyři disky, ale doporučuji nechat alespoň v jedné úrovni (nejlépe 6 nebo 5) aby byla tato důležitá vlastnost RAID polí dostatečně demonstrována.

11.2. Varianta B – Bez výměny sad HDD

Postup v této variantě úlohy bude obdobný jako postup ve variantě A. Zde budou pouze studenti testovat RAID v úrovni, kterou si sestavili v minulé hodině, tzn. nebudou si moci zvolit. Nevýhodou je v této variantě omezení na pouze jednu testovanou úroveň a potencionální možnost, že studenti o připravené pole přijdou dříve, než jej budou moci testovat. Z tohoto důvodu vyžaduje varianta B od studentů větší soustředění na všechny akce při provádění úlohy.

11.3. Vybavení testovacího pracoviště v laboratoři

Pracoviště *viz Obr. 21* obsahuje vlastní zařízení NAS připojené pomocí strukturované kabeláže ke gigabytové počítačové síti (nejlépe vyhrazenému switchi) a druhé zařízení připojené ke stejné síti umožňující čtení a zápis srovnatelnou nebo větší rychlostí než dovoluje NAS.



Obr. 21: Schéma testovacího pracoviště

11.3.1. Použité vybavení

NAS:

Jako NAS zařízení jsem obdržel NAS úložiště N5200B společnosti TheCus *viz Obr. 22.*

Obr. 22: NAS - Thecus N5200B

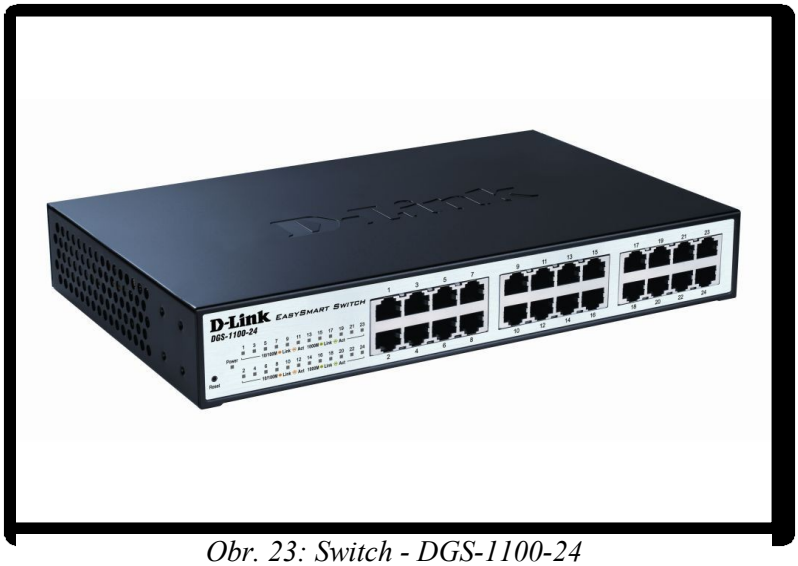
TheCus N5200B je NAS o pěti šachtách pro pevné disky umožňující RAID v úrovních 0, 1, 5, 6, 10 a JBOD. N5200B je určen pro práci v domácí nebo malé kanceláři, nedosahuje tedy parametrů podnikových systémů. Lze u něj například nastavit zálohování ethernetového portu nebo integraci do ACTIVE DIRECTORY struktury a mnoho dalších vlastností známých z rodiny SOHO produktů. Základní parametry NAS systému zobrazuje následující tabulka *Tab. 10.*

Základní parametry Thecus N5200B	
Processor	*Intel® 600MHz Celeron® M*
Memory	*512MB DDR*
WAN	*Gigabit RJ-45 connector*
LAN	*Gigabit RJ-45 connector*
HDD Bays	*5 x 3.5" SATA II HDD, hot-swappable*
HDD Support	*SATA II HDDs up to 1000GB*
eSATA	*1 x eSATA connector for capacity expansion*
USB Ports	*3 x USB type A ports (Host mode) 1 x USB type B ports (Client mode)*

Tab. 10: Základní parametry Thecus N5200B

Gigabit Ethernet Switch:

Jako síťový přepínač byl použit standardní Easy Smart Switch DGS-1100-24 *viz Obr. 23* společnosti D-Link disponující 24 porty, který je možné nahradit prakticky jakýmkoliv gigabytovým switchem na trhu. Přepínač umožňuje použití Jumbo Frame – rozšířený paket pro přenos větších souborů a lepšího využití pásma (v úloze nebyl použit). Tento přepínač umožňuje nastavení pomocí webového prohlížeče přes protokol http (není podmínkou správného průběhu úlohy).



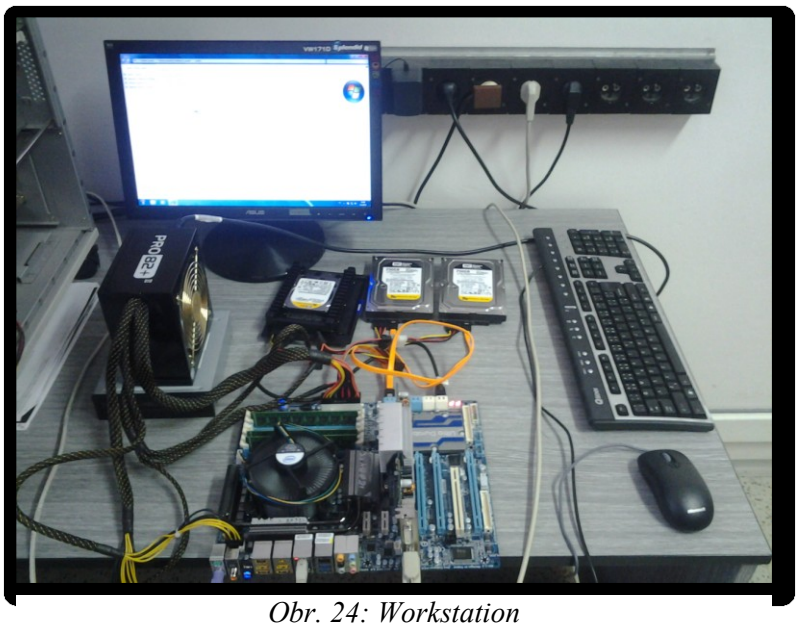
Obr. 23: Switch - DGS-1100-24

Základní parametry D-Link DGS-1100-24	
Rozměry	*11-inch Desktop Size, 1U Height*
Interface	*24 10/100/1000 Mbps ports*
Funkce portu	*IEEE 802.3 compliant* *IEEE 802.3u compliant* *Supports Half/Full-Duplex operation (half at 10/100 Mbps, full at 1Gbps)* *Auto-negotiation* *Auto MDI/MDIX* *IEEE 802.3x Flow Control supports Full-Duplex mode*
Přepínací výkon	*48 Gbps*
Max. Forwarding Rate	*35.71 Mpps*
Velikost tabulky MAC adres	*8K Entries*
Vyrovnávací paměť	*3.5 Mbits*
Flash paměť	*2 MB*

Tab. 11: Základní parametry D-Link DGS-1100-24

Workstation:

Pracovní stanice *viz Obr. 24* byla postavena ze součástek speciálně pro účely testů pole, ale na základě pozdějších měření postačí prakticky každá stanice schopna přehrávat HD video s jedním systémovým diskem a dalším datovým diskem v ideálním případě dvěma disky v RAID 0.



Obr. 24: Workstation

Operační systém je zvolen Microsoft Windows, z důvodu dostupnosti testovacího software Bench32 společnosti ATTO, tudíž testujeme pouze přenos souborů pomocí protokolu SMB/CIFS. Stranou

necháváme protokoly APC a NFS implementované v operačních systémech společnosti Apple resp. systémech Linux.

Parametry stanice uvádím v následující tabulce *Tab. 12.*

Základní parametry Workstation	
Procesor	*INTEL CORE I7-920@2.66GHz*
Základní deska	*Gigabyte GA-X58A-UD5*
Operační paměť	*Kingston KVR1333D3R9NK3/6G*
Grafická karta	*EN7300GT/Silent/256MB*
Operační systém	*Windows 7 64bit CZ Enterprise*
Pevné disky	*1xWestern digital Velociraptor WD4500HLHX* *2x Western digital WD2503ABYX raid edition*
Zdroj	*ENERMAX EPR425APT*

Tab. 12: Workstation - Základní parametry

Strukturovaná kabeláž:

Pro propojení switche, stanice a úložiště byly použity standardní patch kabely kategorie 6 společnosti Roline o délkách 1 a 10m. Při výběru kabelů je nutné brát na zřetel kategorii, pro jakou jsou navrženy v opačném případě tj. při použití kabelu kategorie 5 a nižší budou výsledky laboratorní úlohy zkreslené, jelikož takovýto spoj nevyjedná gigabitový mód a přepne se do stomegabitového módu nebo bude mít nadměrnou chybovost.

11.4. Fáze laboratorní úlohy – Varianta A

Student postupuje po následujících krocích, ve kterých vykonává jednotlivé procedury.

11.4.1. Příprava testovacího pracoviště (cca 5min)

- Vytvoření testovacího pracoviště připojením úložiště a pracovní stanice k síťovému přepínači pomocí patch kabelů
- Uvedení NAS úložiště do výchozího nastavení */viz kap. 11.6.3/*
- Nastavení TCP/IP pracovní stanice pro připojení k NAS */viz kap. 11.6.4/*

11.4.2. Test JBOD (5 disků, cca 10min)

- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Sestavení JBOD */viz kap. 11.6.11/*
- Povolení služby MEDIA SERVER */viz kap. 11.6.7/*
- Připojení sdíleného adresáře a nahrání zkušebních souborů */viz kap. 11.6.8/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/* (výsledek *viz Obr. 27*)
- RESCAN a spuštění VLC přehrávače */viz kap. 11.6.10/*
- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady jednoho disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*

11.4.3. Test RAID 0 (5 disků, cca 10min)

- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Sestavení RAID 0 */viz kap. 11.6.12/*
- Povolení služby MEDIA SERVER */viz kap. 11.6.7/*
- Připojení sdíleného adresáře a nahrání zkušebních souborů */viz kap. 11.6.8/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/* (výsledek *viz Obr. 28*Obr. 27)
- RESCAN a spuštění VLC přehrávače */viz kap. 11.6.10/*
- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady jednoho disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*

11.4.4. Test RAID 1 (2 disky, 1 SPARE disk, cca 15min)

- Vypnutí NAS */viz kap. 11.6.2/*
- Vložení sady disků RAID1
- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Kontrola stavu RAID */viz kap. 11.6.6/*
- Povolení služby MEDIA SERVER */viz kap. 11.6.7/*
- Připojení sdíleného adresáře a nahrání zkušebních souborů */viz kap. 11.6.8/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/*(výsledek *viz Obr. 29*Obr. 27*)*
- RESCAN a spuštění VLC přehrávače */viz kap. 11.6.10/*
- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady jednoho disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*

11.4.5. Test RAID 5 (4 disky, 1 disk SPARE, cca 15min)

- Vypnutí NAS */viz kap. 11.6.2/*
- Vložení sady disků RAID5
- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Kontrola stavu RAID */viz kap. 11.6.6/*
- Povolení služby MEDIA SERVER */viz kap. 11.6.7/*
- Připojení sdíleného adresáře a nahrání zkušebních souborů */viz kap. 11.6.8/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/* (*výsledek viz Obr. 30*Obr. 27*)*
- RESCAN a spuštění VLC přehrávače */viz kap. 11.6.10/*

- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady jednoho disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*

11.4.6. Test RAID 6 – 4 disky, 1disk SPARE (cca. 15min)

- Vypnutí NAS */viz kap. 11.6.2/*
- Vložení sady disků RAID6
- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Kontrola stavu RAID */viz kap. 11.6.6/*
- Povolení služby MEDIA SERVER */viz kap. 11.6.7/*
- Připojení sdíleného adresáře a nahrání zkušebních souborů */viz kap. 11.6.8/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/* (výsledek *viz Obr. 31*Obr. 27)
- RESCAN a spuštění VLC přehrávače */viz kap. 11.6.10/*
- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady jednoho disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/*
- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady druhého disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*

11.4.7. Test RAID 10 – 4 disky, 1 SPARE disk (cca 15min)

- Vypnutí NAS */viz kap. 11.6.2/*
- Vložení sady disků RAID0
- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Kontrola stavu RAID */viz kap. 11.6.6/*
- Povolení služby MEDIA SERVER */viz kap. 11.6.7/*
- Připojení sdíleného adresáře a nahrání zkušebních souborů */viz kap. 11.6.8/*
- Testování nástrojem Bench32 */viz kap. 11.6.9/*(výsledek *viz Obr. 32)*
- RESCAN a spuštění VLC přehrávače */viz kap. 11.6.10/*
- Vypnutí NAS (rozhodne vyučující) */viz kap. 11.6.2/*
- Simulace závady jednoho disku
- Zapnutí NAS */viz kap. 11.6.1/*
- Zhodnocení stavu a funkce NAS */viz kap. 11.6.14/*

11.5. Fáze laboratorní úlohy – Varianta B

11.5.1. Příprava testovacího pracoviště (cca 5min.)

- Postup shodný s kapitolou11.4.1

11.5.2. Test připravené úrovně RAID 1 resp. 5 resp. 6 resp. 10 (cca 15min.)

- Postup shodný s kapitolou 11.4.4 resp. 11.4.5 resp. 11.4.6 resp. 11.4.7(bez vkládání HDD)

11.5.3. Test JBOD (5 disků, cca 10min)

- Postup shodný s kapitolou 11.4.2

11.5.4. Test RAID 0 (5 disků, cca 10min)

- Postup shodný s kapitolou 11.4.3

11.5.5. Sestavení pole RAID 1 resp. 5 resp. 6 resp. 10 (5 disků, cca 5min.)

- Zapnutí NAS */viz kap. 11.6.1/*
- Přihlášení do NAS jako uživatel admin */viz kap. 11.6.5/*
- Sestavení pole RAID 1 resp. 5 resp. 6 resp. 10 (disk 1,2,3,4,SPARE) */viz kap. 11.6.1311.6.6/*

11.6. Procedury

11.6.1. Zapnutí NAS

- Zapnutí NAS provedeme stiskem tlačítka POWER na předním panelu.

11.6.2. Vypnutí NAS

Lze provést dvojím způsobem buď prostřednictvím hardwaru, nebo softwaru:

- Prostřednictvím hardwaru lze vypnout zařízení stlačením tlačítka POWER na předním panelu.
- Prostřednictvím softwaru vypneme zařízení tak, že po přihlášení jako admin (viz.kapitola11.6.5) v hlavním menu zvolíme SYSTÉM a z podmenu vybereme položku REBOOT & SHUTDOWN. Na vyobrazené stránce aktivujeme volbu SHUTDOWN.

11.6.3. Uvedení NAS úložiště do výchozího nastavení

- Po úspěšném připojení napájení propojíme patch kabelem WAN port úložiště se sítí ethernet.
- Uvedeme zařízení do chodu pomocí síťového vypínače
- Po skončení bootovací procedury se ozve tón, který oznamuje, že zařízení je připraveno k použití.
- Na čelním panelu stiskneme tlačítko ENTER a zadáme heslo (defaultně: 0000)
- Po úspěšném zadání hesla se zobrazí první položka menu a my tlačítky nahoru resp. dolů vybereme položku RESET TO DEFAULT, potvrdíme

tlačítkem ENTER a ještě jednou potvrdíme výběrem YES a stisknutím ENTER.

- Tím je nastavení do výchozího stavu dokončeno a po dokončení startovací procedury a po zvukovém signálu bude NAS připraveno k další konfiguraci přes http prohlížeč.

11.6.4. Nastavení TCP/IP pracovní stanice pro připojení k NAS

Je třeba nastavit TCP/IP protokol, u obou zařízení tak, aby byly na stejné podsíti (doporučuje se ponechat defaultní nastavení NAS po resetu tj. 192.168.1.100) tak jak je uvedeno na schématu *Obr. 21*

Všechny následující postupy vycházejí z předpokladu, že je použita česká verze operačního systému a připojení k síti nese název PŘIPOJENÍ K MÍSTNÍ SÍTI:

IP ADRESA: 192.168.1.200 MASKA: 255.255.255.0 BRÁNA: 192.168.1.200

Citované parametry lze nastavit pomocí grafického rozhraní nebo pomocí příkazového řádku:

- Windows XP: Pomocí myši zvolíme START->SPUSTIT a do příkazové řádky napíšeme CMD a zvolíme OK. Do spuštěného okna pak vepíšeme příkaz: NETSH INTERFACE IP SET ADDRESS NAME="PŘIPOJENÍ K MÍSTNÍ SÍTI" STATIC 192.168.1.200 255.255.255.0 192.168.1.200
- Windows Vista / Windows 7: Pomocí myši zvolíme START a do vyhledávacího pole napíšeme CMD a po vyhledání v horní části dialogového okna klikneme na program pravým tlačítkem myši a zvolíme SPUSTIT JAKO SPRÁVCE. Do spuštěného okna pak vepíšeme příkaz: NETSH INTERFACE IPV4 SET ADDRESS NAME="PŘIPOJENÍ K MÍSTNÍ SÍTI" STATIC 192.168.1.200 255.255.255.0 192.168.1.200

11.6.5. Přihlášení do NAS jako uživatel „admin"

- Spustíme http prohlížeč a zadáme do adresního řádku IP ADRESU NAS úložiště, pro IP:192.168.1.100:

  http://192.168.1.100

- Po načtení přihlašovací stránky se přihlásíme pod tímto účtem:

USERNAME: admin

PASSWORD: admin

- Pokud se přihlašujeme do úložiště poprvé, je nutné přijmout zobrazené prohlášení zatržením checkboxu a zvolením APPLY, v opačném případě se hned po zvolení tlačítka LOGIN dostáváme na stránku s informacemi o stavu systému a tím jsme se úspěšně přihlásili do administrátorského prostředí.

11.6.6. Kontrola stavu RAID

- Pod přihlášeným uživatelem admin zvolíme z hlavního menu STORAGE a podpoložku RAID
- Na vyobrazené stránce zkontrolujeme úroveň RAID, a zda se pole nachází ve stavu HEALTHY tj. zda je „zdravé“.

11.6.7. Povolení služby MEDIA SERVER

- Pod přihlášeným uživatelem admin vybereme z hlavního menu NETWORK podpoložku MEDIA SERVER.
- Na zobrazené stránce zkontrolujeme, že radiobutton je v poloze DISABLED.
- Pokud tak není, přepneme jej do polohy DISABLED a aktivujem volbu pomocí tlačítka APPLY.
- Nyní již můžeme službu opětovně povolit přepnutím radiobuttonu do polohy ENABLE a následnou aktivací tlačítka APPLY.
- Povolením nebo opětovným povolením služby jsme si zajistili automatické vytvoření adresáře MY MEDIA na RAID poli, také automatické přidělení přístupového oprávnění PUBLIC a konečně i automatické nastavení sdílení. Tuto akci musíme ještě povolit aktivací tlačítka YES a přijmutím oznámení o průběhu vytváření adresáře po výběru tlačítka APPLY z předchozího bodu.

- Adresář MYMEDIA je ještě třeba přidat do seznamu skenovaných adresářů, což učiníme zaškrtnutím checkboxu na téže stránce a potvrdíme oznámení o úspěšném přidání.

11.6.8. Připojení sdíleného adresáře a nahrání zkušebních souborů

- Na pracovní stanici s operačním systémem Microsoft Windows 7 / Vista zvolíme pomocí myši START a v pravém podmenu aktivujeme tlačítko SÍŤ. V otevřeném okně zvolíme server s názvem N5200. V následujícím okně aktivací pravého tlačítka myši na adresáři MY MEDIA zpřístupníme volbu PŘIPOJIT SÍŤOVOU JEDNOTKU, kterou aktivujeme. V otevřeném dialogovém okně vybereme zástupné písmeno X pro síťovou jednotku a zvolíme DOKONČIT.

- Opět volíme START a v pravém podmenu zpřístupníme pravým tlačítkem myši na položce POČÍTAČ volbu OTEVŘÍT. V otevřeném průzkumníku zkopírujeme zkušební soubory z daného umístění na sdílený adresář NAS zařízení reprezentovaný sdílenou jednotkou pod písmenem X.

- Na pracovní stanici s operačním systémem Microsoft Windows XP zvolíme pomocí myši START a v pravém podmenu aktivujeme tlačítko TENTO POČÍITAČ. V otevřeném okně zvolíme z horního menu položku SLOŽKY a v tomto stromu rozevřeme větev MÍSTA V SÍŤI. Dále rozevřeme větev CELÁ SÍŤ a vybereme položku SÍŤ MICROSOFT WINDOWS. V pravém panelu by se po chvíli měli objevit skupiny, které obsahuje konkrétní síť a mezi nimi také defaultní skupina úložiště MYGROUP, jež obsahuje i naše úložiště N5200. Po otevření NAS se zobrazí sdílené adresáře, které obsahují mimo jiné adresář MY MEDIA. Na adresáři aktivujeme pravé tlačítko myši a zvolíme PŘIPOJIT SÍŤOVOU JEDNOTKU. V otevřeném průvodci změníme pomocí roletového seznamu zástupné písmeno sdíleného síťového disku X.

- Po úspěšném připojení síťové jednotky zkopírujeme v již otevřeném průzkumníkovi testovací soubory z daného umístění na jednotku X

11.6.9. Testování nástrojem Bench32

- Na pracovní stanici zvolíme START a do vyhledávacího pole (Windows 7) resp. po výběru položky SPUSTIT (Windows XP) vepíšeme text EXPLORER a potvrdíme klávesou ENTER. V otevřeném okně průzkumníku spustíme dvojklikem z daného umístění soubor BENCH32.EXE
- V okně programu BENCH32 změníme parametr s názvem DRIVE: tak, že vybereme z roletového seznamu písmeno síťové jednotky X náležící sdílenému adresáři v NAS úložišti.
- Ostatní parametry ponecháme nezměněny a spustíme test aktivací tlačítka START.
- Naměřené hodnoty zaneseme do protokolu.

11.6.10. RESCAN a spuštění VLC přehrávače

- V konfiguračním prostředí NAS úložiště a přihlášení pod uživatelem admin zvolíme z hlavního menu NETWORK a z podnabídky zvolíme položku MEDIA SERVER.
- Na stránce nastavení MEDIA SERVERu aktivujeme tlačítko RESCAN u adresáře MY MEDIA. Touto volbou jsme aktualizovali seznam stop, nebo chce-li souborů, které server „vysílá“ do sítě, o čemž po ukončení aktualizace zobrazí zprávu.
- Nyní již můžeme na pracovní stanici spustit z daného umístění VLC přehrávač.
- Současným stiskem kláves CTRL a L se přepneme do seznamu skladeb programu VLC. Zde pomocí myši rozevřeme podmenu u položky MÍSTNÍ SÍŤ a vybereme UNIVERSAL PNP. Pochvíli by se měl objevit N5200 multimediální server se strukturovaným seznamem stop,

ze kterého lze vybrat konkrétní stopu a spustit tlačítkem PLAY. Tím je funkce MEDIA SERVERU ověřena.

11.6.11. Sestavení pole JBOD (disk 1,2,3,4,5)

- V administrátorském prostředí zvolíme z hlavního menu STORAGE, dále pak DISKS a zkontrolujeme, zda systém obsahuje všech pět disků a zda jsou ve stavu OK.
- Zvolíme z hlavního menu opět STORAGE a z podmenu RAID.
- Pokud je aktuálně pole jakékoliv úrovně sestaveno, je na stránce RAID INFORMATION vyobrazeno a tlačítko NEW je neaktivní. V tomto případě je nutné aktuální pole odstranit volbou tlačítka RAID CONFIG následované volbou REMOVE RAID, která si vyžádá v dalším kroku potvrzení, tj. opsání textu Yes (záleží na velikosti písmen!).
- Nyní již můžeme na stránce RAID INFORMATION použít volbu NEW a pole typu JBOD nakonfigurovat.
- Zvolíme všechny disky z výběru, parametr STRIPE SIZE ponecháme defaultní tj. 64KB a parametrem USED PERCENTAGE, který přímo ovlivňuje dobu konstrukce pole, omezíme velikost pole na 1% (tj. cca 22GB při použití disků o kapacitě 500GB)
- Aktivujeme tlačítko CREATE a potvrdíme
- Proběhne konstrukce a formátování pole (cca 4minuty)

11.6.12. Sestavení pole RAID 0 (disk 1,2,3,4,5)

- V administrátorském prostředí zvolíme z hlavního menu STORAGE, dále pak DISKS a zkontrolujeme, zda systém obsahuje všech pět disků a zda jsou ve stavu OK.
- Zvolíme z hlavního menu opět STORAGE a z podmenu RAID.
- Pokud je aktuálně pole jakékoliv úrovně sestaveno, je na stránce RAID INFORMATION vyobrazeno a tlačítko NEW je neaktivní. V tomto případě je nutné aktuální pole odstranit volbou tlačítka RAID CONFIG

následované volbou REMOVE RAID, která si vyžádá v dalším kroku potvrzení, tj. opsání textu Yes (záleží na velikosti písmen!).

- Nyní již můžeme na stránce RAID INFORMATION použít volbu NEW a pole RAID 0 nakonfigurovat.
- Zvolíme všechny disky z výběru, parametr STRIPE SIZE ponecháme defaultní tj. 64KB a parametrem USED PERCENTAGE, který přímo ovlivňuje dobu konstrukce pole, omezíme velikost pole na 1% (tj. cca 22GB při použití disků o kapacitě 500GB)
- Aktivujeme tlačítko CREATE a potvrdíme
- Proběhne konstrukce a formátování pole (cca 3minuty)
- Proběhne konstrukce a formátování pole (cca 4minuty)

11.6.13. Sestavení pole RAID 1 resp. 5 resp. 6 resp. 10 (disk 1,2,3,4,SPARE)

- V administrátorském prostředí zvolíme z hlavního menu STORAGE, dále pak DISKS a zkontrolujeme, zda systém obsahuje všech pět disků a zda jsou ve stavu OK.
- Zvolíme z hlavního menu opět STORAGE a z podmenu RAID.
- Pokud je aktuálně pole jakékoliv úrovně sestaveno, je na stránce RAID INFORMATION vyobrazeno a tlačítko NEW je neaktivní. V tomto případě je nutné aktuální pole odstranit volbou tlačítka RAID CONFIG následované volbou REMOVE RAID, která si vyžádá v dalším kroku potvrzení, tj. opsání textu Yes (záleží na velikosti písmen!).
- Nyní již můžeme na stránce RAID INFORMATION použít volbu NEW a pole RAID požadované úrovně nakonfigurovat.
- Zvolíme všechny disky z výběru kromě jednoho, u kterého zaškrtneme SPARE, parametr STRIPE SIZE ponecháme defaultní tj. 64KB a stejně tak i parametr USED PERCENTAGE, který omezuje velikost pole, ponechme nezměněn.
- Aktivujeme tlačítko CREATE a potvrdíme

- Proběhne konstrukce, formátování a sestavení pole (max. 500min.)

11.6.14. Zhodnocení stavu a funkce NAS

- Po kompletním startu systému se přihlásíme do systému jako admin (viz procedura 11.6.5) vybereme z hlavního menu položku STORAGE a z podmenu RAID. Na stránce informací o sestaveném RAID poli odečteme stav pole.
- Na RAID poli po simulaci chyby jednoho disku spustíme, pokud to stav pole umožňuje a sdílená jednotka je dostupná, testování nástrojem Bench32 (viz procedura 11.6.9).
- Spustíme VLC přehrávač a zhodnotíme stav a kvalitu přehrávání zvláště u stop s vysokým nárokem na propustnost tj. především u videa s vysokým rozlišením.
- V průběhu inicializace pole – BUILDING, čtení dat a zápisu dat pozorujeme aktivitu jednotlivých disků pole podle jejich stavových LED (aktivní současně/ jeden po druhém apod.).
- Výsledky jednotlivých bodů zaznamenáme do protokolu.

12. Závěr

12.1. Teoretická část

Přiblížil a vysvětlil jsem čtenářovi architekturu RAID systémů, nastínil okolnosti vzniku a podíval se, pro lepší přehled čtenáře částečně, i do historie úložišť dat.

Dříve než jsem se začal zaobírat filosofií samotných polí, rozebral jsem a rozdělil úložiště obecně, vysvětlil jsem rozdíl mezi úložišti typu NAS, DAS a SAN. Podrobně uvedl rozdílná rozhraní a technologie, se kterými se může vážený čtenář setkat.

Sumarizoval jsem a popsal všechny nejpoužívanější úrovně, uvedl u nich důležité parametry jako například spolehlivost, výkon a velikost režie. V textu jsem vypíchnul diskutabilní vlastnosti všech úrovní a jejich klady a zápory. Současně jsem uvedl a popsal schematicky jejich princip a fyzické rozprostření ukládaných dat na jednotlivých členech pole. Důležité parametry dvou hlavních skupin jsem uvedl v přehledných tabulkách na závěr každé stati.

Uvedl jsem technologie, které jsou dle mého názoru stěžejní s ohledem na možný rozvoj, který by si jistě zasloužili. Jsou to zejména SSD disky a možnosti jejich paměťových buněk, které nejsou v současné době nikterak oslnivé, takže je zde prostor pro vývoj. Ze softwarových produktů čeká jistě masivnější rozšíření ZFS a jeho architektura RAID Z, jejíž vývoj je ve stádiu možného nasazení, avšak doposud není tak rozšířen, jak by mohl a měl být.

V této souvislosti si dovolím navrhnout téma na nadstavbu této práce, která by se soustředila pouze na konstrukci, výkonnostní a spolehlivostní parametry úložiště postaveném na základu RAID Z2 a osazeném SSD disky. Operační systém nechť laskavě zvolí sám autor, jelikož v současné době není portace na operační systém Linux ještě dokončena a je k dispozici pouze verze pro Open Solaris a FreeBSD (pro některé další systémy jsou k dispozici pouze starší verze).

12.2. Flash aplikace

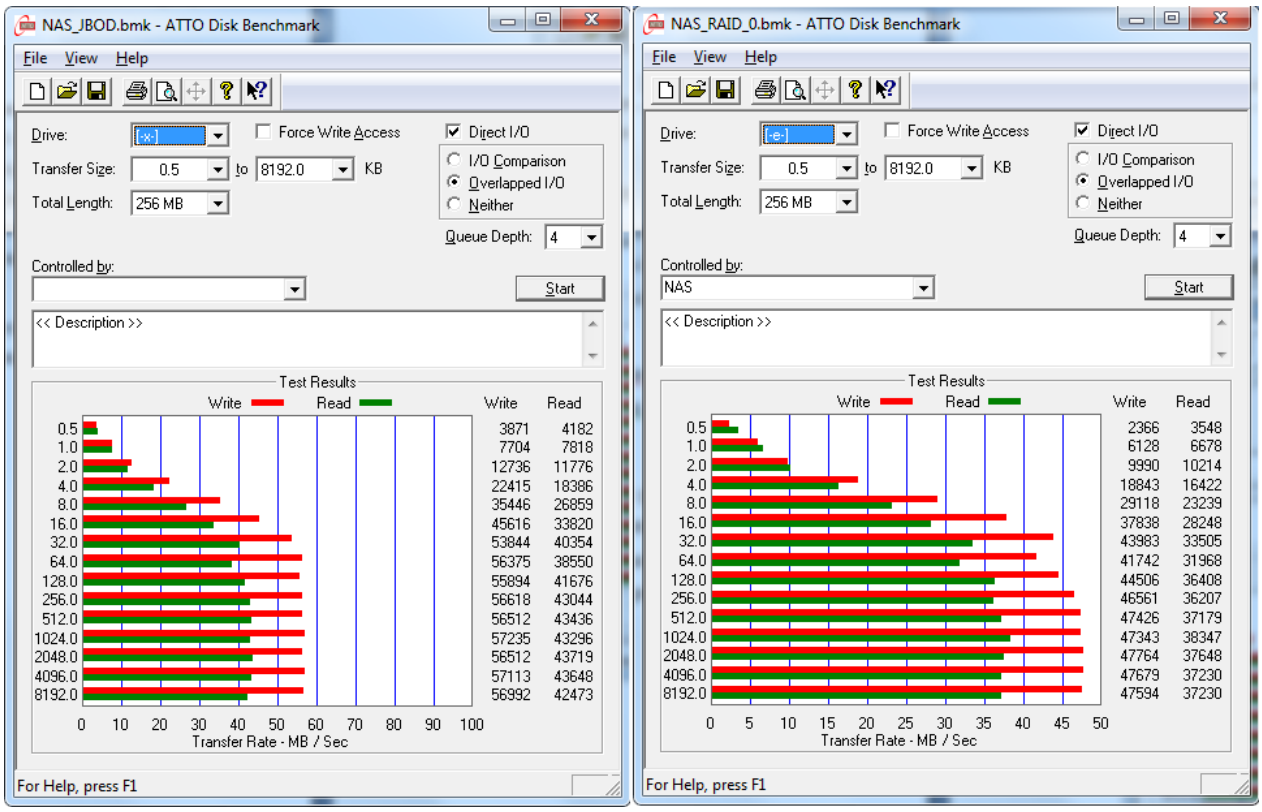
Flash aplikace obsahuje dvě části a to teoretickou a praktickou. V teoretické části student obdrží základní informace o RAID architektuře prostřednictvím interaktivních slidů. V praktické části nabyté informace uplatní při laboratorní úloze, kterou ho interaktivní průvodce provede.

Všechna použitá tlačítka jsem obsloužil pomocí metody object.addEventListener (MouseEvent.CLICK,fl_funkcion) a dále voláním podmíněných skoků na jednotlivé kapitoly *viz. Obr. 33.*

12.3. Návrh laboratorní úlohy

Úloha ukazuje studentům vlastnosti a využití redundantních polí v praxi, ovšem s určitými omezeními danými časovým limitem a použitým NAS systémem. Sestavení pole jednotlivých úrovní, představuje časy uvedené v tabulce *Tab. 9*. Tyto jsou neakceptovatelné pro demonstraci několika úrovní v jedné vyučovací hodině. Řešením je pole sestavit v předchozí hodině a výsledné pole otestovat po příslušné době v hodině následující nebo disponovat několika sadami disků s již sestavenými poli. Blíže se touto tématikou zaobírám v těle dokumentu v příslušné kapitole.

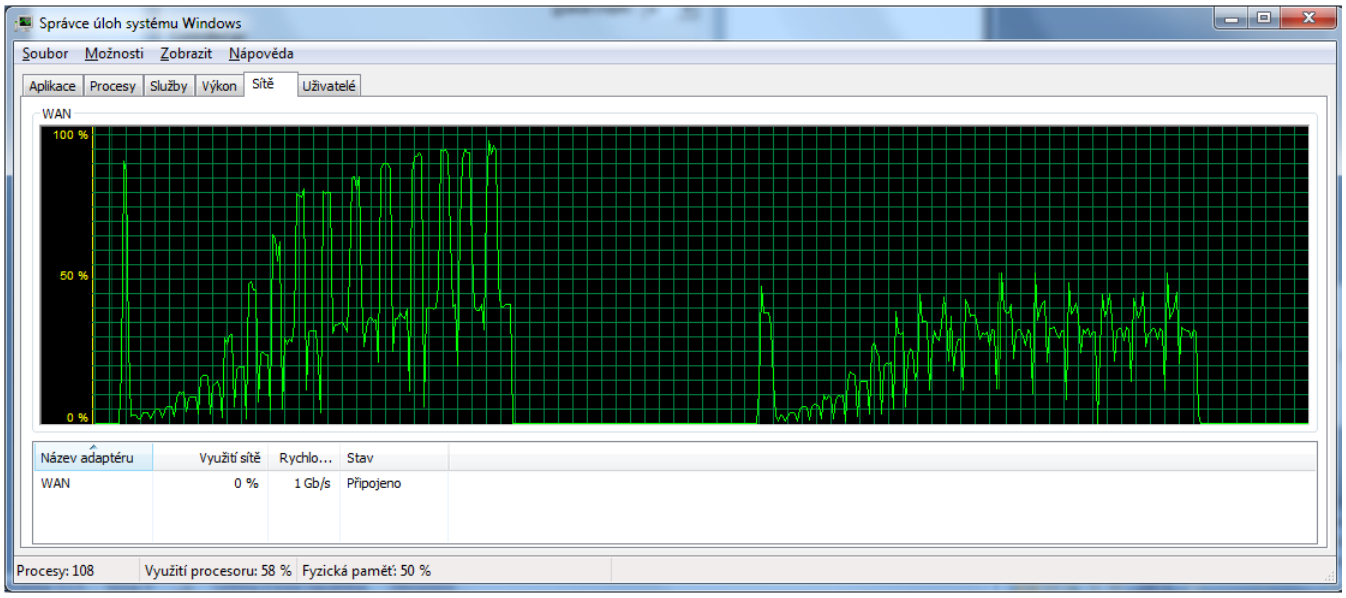
Při testování propustnosti na zapůjčeném zařízení Thecus N5200B bylo dosaženo maximálního datového toku na poli JBOD a nikoliv na RAID 0, kde by teoreticky měla být propustnost nejvyšší.



Obr. 25: NAS - Porovnání výkonu RAID0/JBOD

Tvrzení dokazují naměřené hodnoty výše *viz Obr. 25(vlevo JBOD, vpravo RAID 0/poloviční měřítko na ose X/)*. Měření bylo provedeno dvěma nástroji – Bench32 společnosti ATTO a NAS performance toolkit společnosti Intel s totožnými výsledky.

V této souvislosti jsem provedl rozsáhlé diagnostikování, zda se nevyskytla chyba měření. Jako příklad mohu uvést analýzu síťového přenosu *viz Obr. 26vpravo – NAS RAID 0)*.



Obr. 26: Analýza síťového provozu (vlevo PC-RAID0, vpravo NAS-RAID0)

Dále bylo zkoumáno zdrojové, resp. cílové úložiště, které tvořilo protistranu při měřeních, zda je schopno přijímat resp. vysílat data požadovanou rychlostí. Měřením bylo zjištěno, že softwarové pole RAID 0 sestavené na standardní stanici platformy Intel s OS Microsoft Windows 7 *viz Obr. 26(vlevo – PC RAID 0)* je schopno přenášet data rychlostí, jež je limitována až propustností sítě. Dalšími provedenými měřeními ve shodě s *lit. [17]* jsem potvrdil, že uvedený systém není schopen přenášet data větší nežli změřenou rychlostí. Což se negativně projeví na názornosti úlohy, nicméně i přes tyto nedostatky je cvičení smyslupnlné.

13. Použitá literatura

[1] Patterson,David A., Gibson GARTH, Randy H. KATZ. A Case for Redundant Arrays of Inexpensive Disks (RAID) [online]. 1987 [cit. 2012-02-10]. Dostupné z: <http://www.eecs.berkeley.edu/Pubs/TechRpts/1987/CSD-87-391.pdf>

[2] AMDAHL,G.M. Validity of the single procesor approach to achieving large scale computing capabilities [online]. 1967 [cit. 2012-02-10]. Dostupné z: <http://people.cs.umass.edu/~emery/classes/cmpsci691st/readings/Conc/Amdahl-04785615.pdf>

[3] GUPTA, Meeta Storage Area Network Fundamentals.Indianapolis:Cisco Press, 2002.ISBN 978-1-58705-065-7

[4] INCITS. International Committee for Information Technology Standards [online]. 2012. Dostupné z:< http://www.incits.org/>

[5] SCSITA. In:SCSI Trade Association [online]. 2011 [cit. 2012-02-10]. Dostupné z: <http://www.scsita.org/terms-and-terminology.html>

[6] SCSI. In: Wikipedia:The free encyclopedia [online]. [2006], 10.04.2012 [cit. 2012-02-07]. Dostupné z: <http://en.wikipedia.org/wiki/SCSI>

[7] Parallel ATA. In: Wikipedia:The free encyclopedia [online]. [2006], 17.04.2012 [cit. 2012-02-07]. Dostupné z: <http://en.wikipedia.org/wiki/Parallel_ATA>

[8] Serial ATA. In: Wikipedia:The free encyclopedia [online]. [2006], 17.04.2012 [cit. 2012-02-07]. Dostupné z: <http://en.wikipedia.org/wiki/Serial_ATA>

[9] USM. In:SATA International Organization [online]. [2010], [cit. 2012-02-12]. Dostupné z: <http://www.sata-io.org/documents/SATA-IO-Whitepaper_USM.pdf>

[10] Fibre channel. In: Wikipedia:The free encyclopedia [online]. [2006], 28.04.2012 [cit. 2012-02-07]. Dostupné z: <http://en.wikipedia.org/wiki/Fibre_Channel>

[11] BAARF. Battle Against Any RAID Five [online]. [2003]. Dostupné z: <http://www.baarf.com/>

[12] Leventhal, Adam. Triple – Parity RAID and Beyond [online]. [2009], [cit. 2012-01-12]. Dostupné z: < http://queue.acm.org/detail.cfm?id=1670144>

[13] RAID. In: Wikipedia:The free encyclopedia [online]. [2006], 02.05.2012 [cit. 2012-02-17]. Dostupné z: <http://en.wikipedia.org/wiki/RAID>

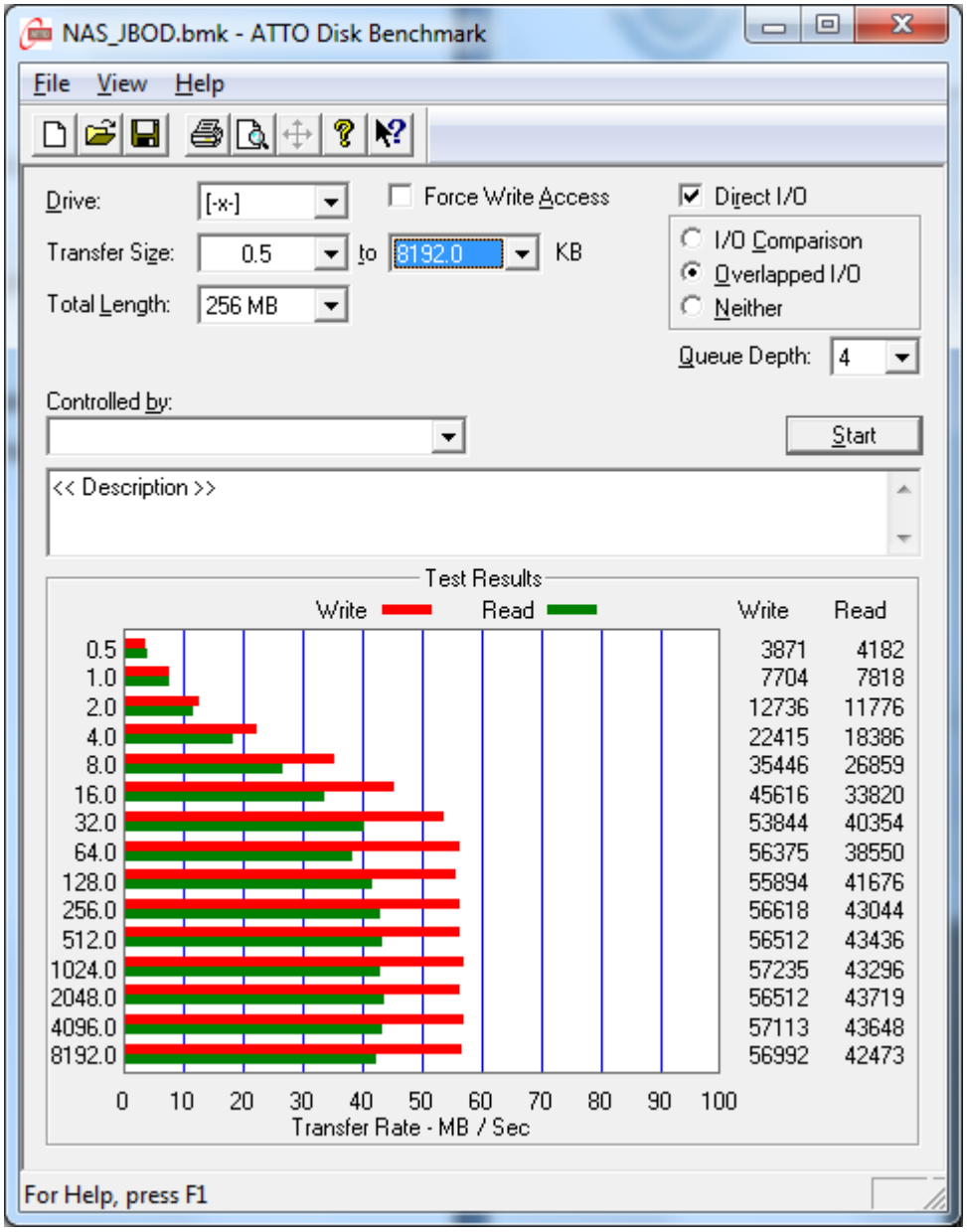
[14]] Nested RAID levels. In: Wikipedia:The free encyclopedia [online]. [2006], 02.05.2012 [cit. 2012-02-17]. Dostupné z: <http://en.wikipedia.org/wiki/Nested_RAID_levels>

[15] Intel® Matrix Storage Technology. In: Intel® Corporation [online]. [cit. 2012-02-17]. Dostupné z: <http://www.intel.com/design/chipsets/matrixstorage_sb.htm>

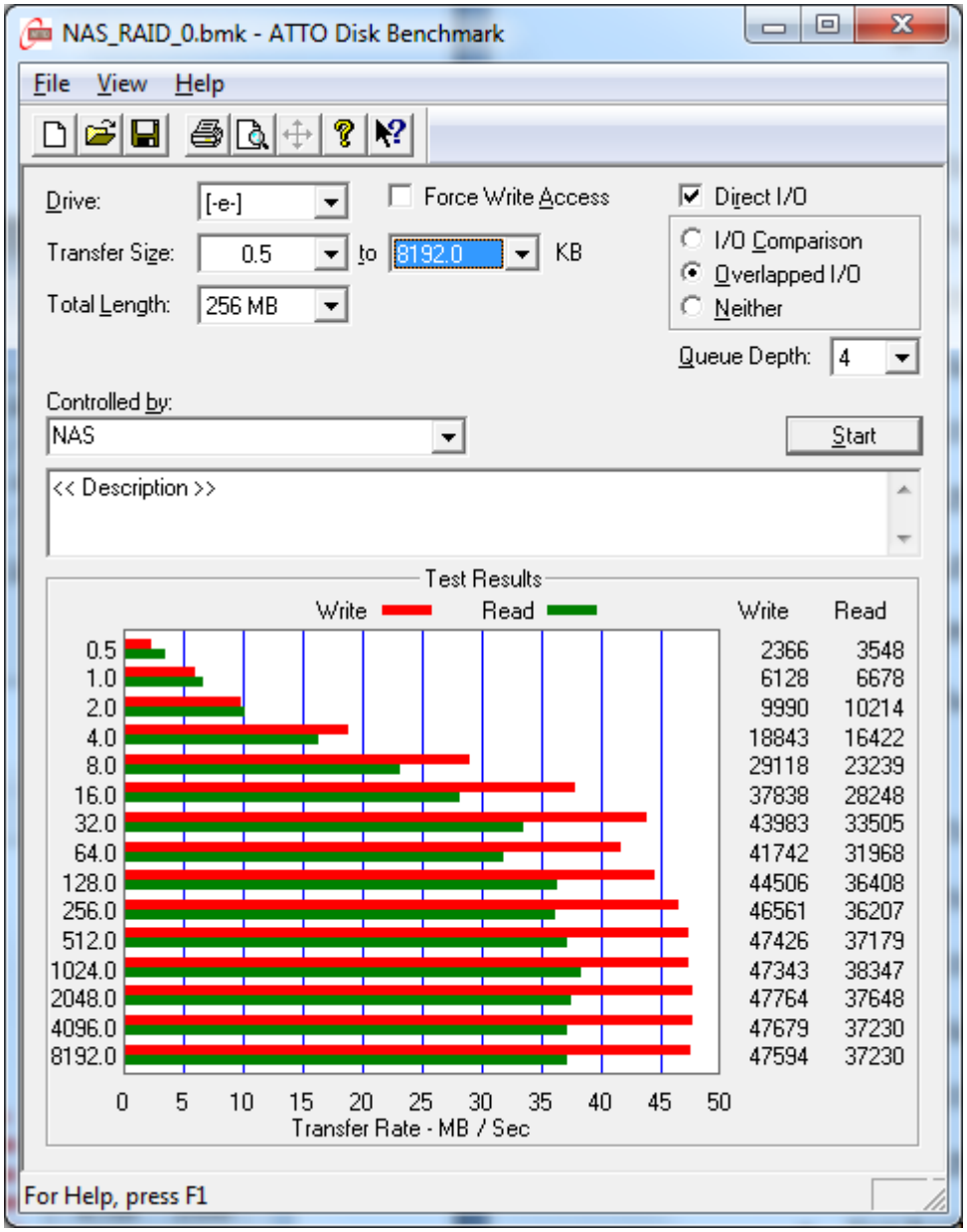
[16] Panzer-Steindel, Bernd. Data integrity [online]. [2007], [cit. 2012-02-23]. Dostupné z: <http://indico.cern.ch/getFile.py/access?contribId=3&sessionId=0&resId=1&materialId=paper&confId=13797>

[17] Copy Files to NAS. In: Tom's Hardware [online]. [2012], [cit. 2012-02-17]. Dostupné z: <http://www.tomshardware.com/charts/network-attached-storage-nas-charts/Copy-Files-to-NAS,1917.html>

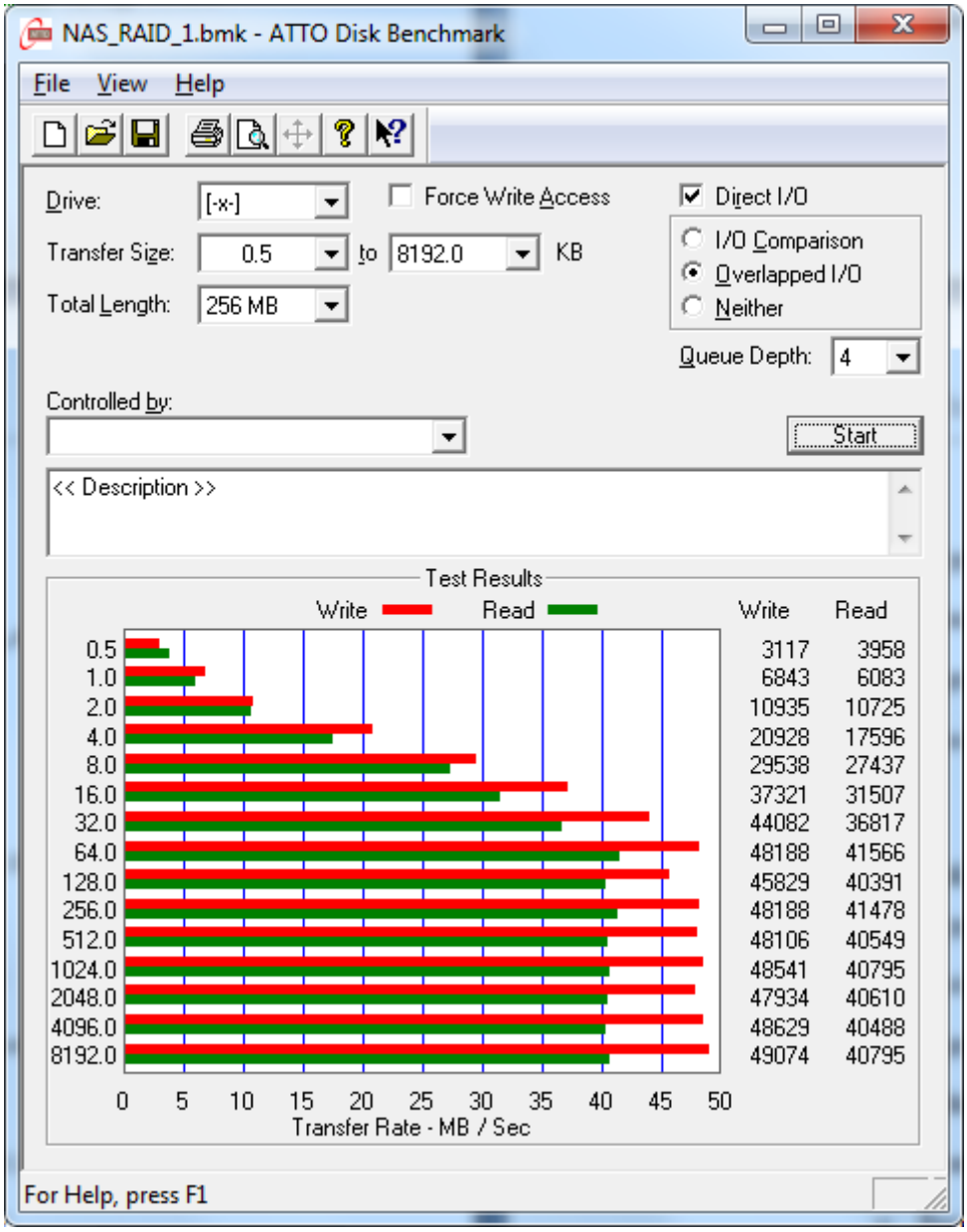
14. Seznam příloh:



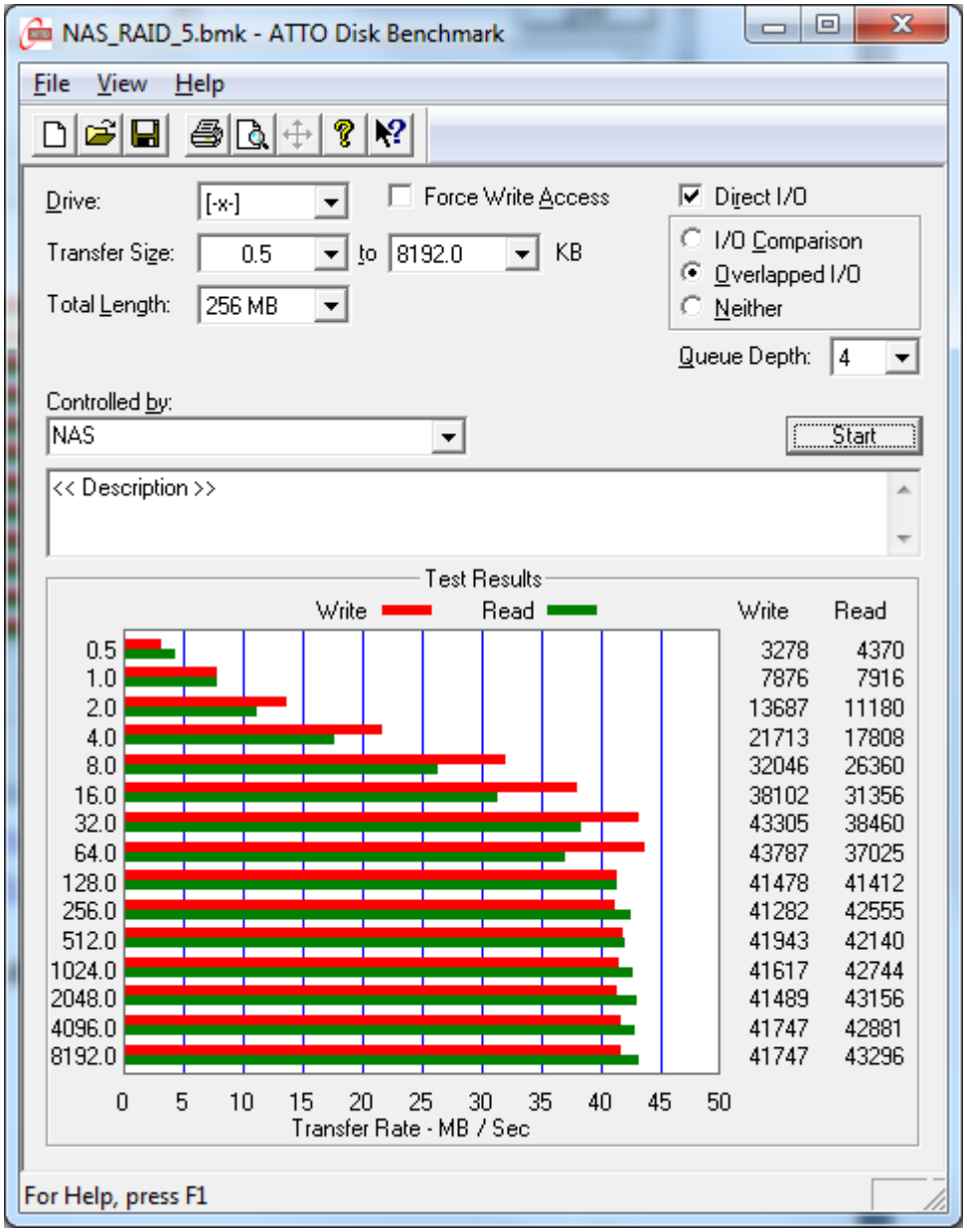
Obr. 27:Měření propustnosti JBOD



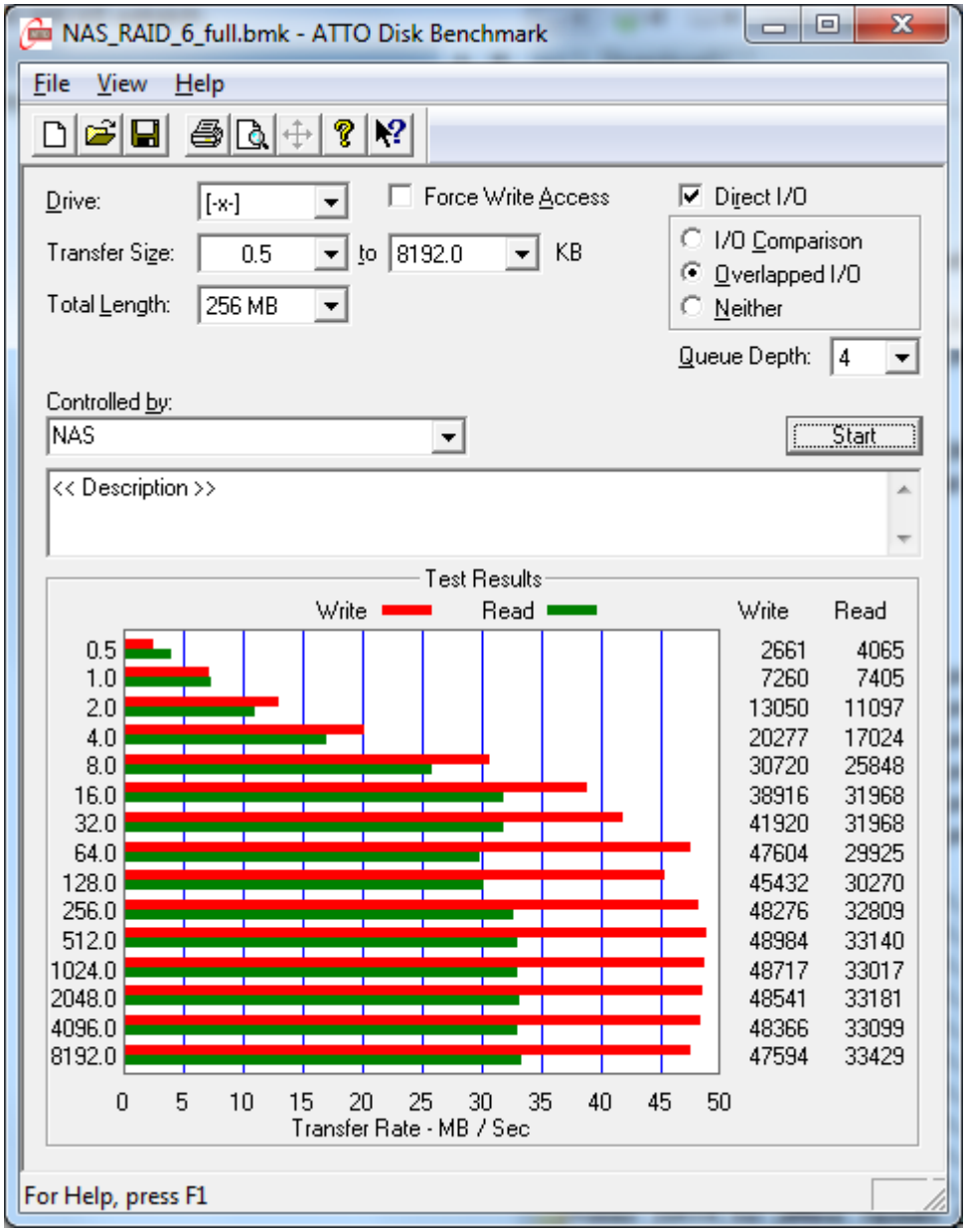
Obr. 28: Měření propustnosti RAID 0



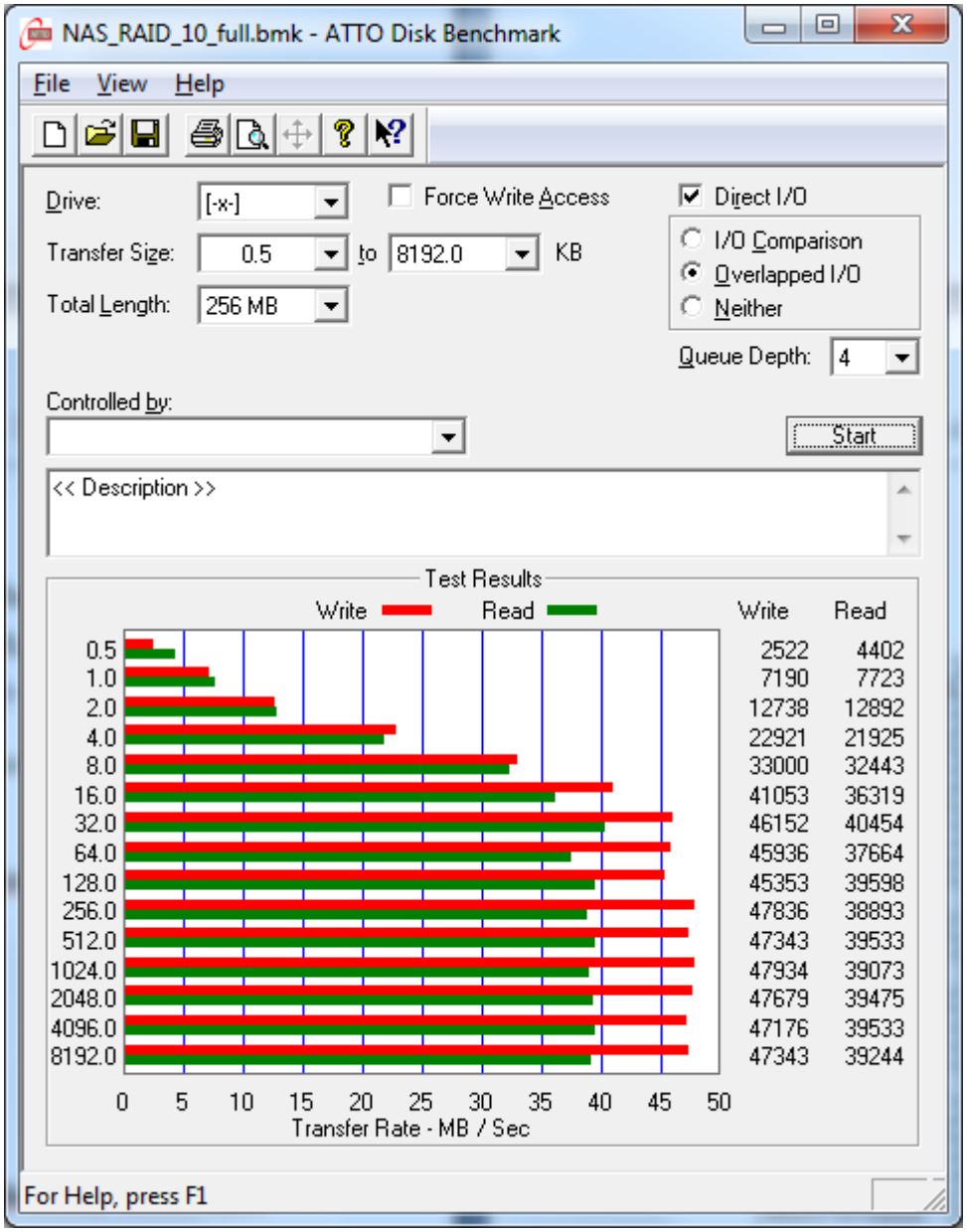
Obr. 29: Měření propustnosti RAID 1



Obr. 30: Měření propustnosti RAID 5



Obr. 31: Měření propustnosti RAID 6



Obr. 32: Měření propustnosti RAID 10

```
import flash.events.MouseEvent;

//Navigace
btn_Spustit.addEventListener(MouseEvent.CLICK, fn_Dale);

//Definice funkcí
function fn_Zpet(event:MouseEvent)
{
//  trace("pred:"+currentFrame);
    prevFrame();
//  trace("po:"+currentFrame);
};

function fn_Dale(event:MouseEvent)
{
//  trace("pred:"+currentFrame);
    nextFrame();
//  trace("po:"+currentFrame);
};

function fn_JdiNekam(Stitek:String)
{
    gotoAndStop(Stitek);
};


stop();
btn_Dale.addEventListener(MouseEvent.CLICK, fn_Dale);
btn_Zpet.addEventListener(MouseEvent.CLICK, fn_Zpet);
btn_Obsah.addEventListener(MouseEvent.CLICK, function(){fn_JdiNekam("Rozcestnik");});

//Dalsi tlacitka
tlf_ArchRAIDoko.addEventListener(MouseEvent.CLICK, function(){fn_JdiNekam("sl_ArchRAIDokol");});
tlf_ArchRAIDprin.addEventListener(MouseEvent.CLICK, function(){fn_JdiNekam("sl_ArchRAIDprin");});

stop();
//trace(currentFrame);
```

Obr. 33: Action script 3.0

Bench_32	Adresář obsahující testovací software
Test_Files	Adresář obsahující testovací soubory
VLC_Portable	Adresář obsahující přehrávač VLC
PaligaJan-DP_2012_final	Diplomová práce ve formátu Microsoft Word 2007 (.docx)
PaligaJan-DP_2012_final	Diplomová práce ve formátu PDF/A (.pdf)
RAID	FLASH aplikace ve formátu spustitelného souboru pro win32 (.exe)
RAID	FLASH zdrojový kód pro Adobe Flash CS5 (.fla)
RAID	HTML zavaděč FLASH aplikace
RAID	FLASH aplikace (.swf)

Tab. 13: Obsah DVD